A Dissertation

on

Study on Opinion Leader in Online Social Network using Computational Intelligence

Submitted in fulfilment of the requirements
for the award of the Degree of

Doctor of Philosophy

By

LOKESH JAIN
(2K17/PhD/CO/10)

Under the Joint Supervision of

Prof. Rahul Katarya (Supervisor)
Department of CSE, DTU, Delhi

Dr. Shelly Sachdeva (Co-Supervisor)
Associate Prof. & HOD, Department of CSE, NIT, Delhi

To

Department of Computer Science and Engineering

Delhi Technological University, Delhi

2021

CANDIDATE DECLARATION

I hereby declare that the thesis entitled ***"Study on Opinion Leader in Online Social Network using Computational Intelligence"*** submitted to Delhi Technological University, Delhi in the partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy in the Department of Computer Science, is an authentic record of my work under the supervision of Prof. Rahul Katarya (Supervisor), Department of Computer Science & Engineering, DTU, Delhi and Dr. Shelly Sachdeva (Co-Supervisor), Associate Professor and HOD, Department of Computer Science & Engineering, NIT, Delhi.

The interpretations put forth are based on my reading and understanding of the original texts, and they are not published anywhere in the form of books, monographs, or articles.

The matter embodied in this thesis has not been submitted by me for the award of any other degree.

Lokesh Jain
Roll no: 2K17/PhD/CO/10
Department of Computer Science
Delhi Technological University
Delhi, India-110042

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

(Govt. of National Capital Territory of Delhi)

Shahbad Daulatpur, Main Bawana Road,
Delhi-110042. India

Date: _______________

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled **"Study on Opinion Leader in Online Social Network using Computational Intelligence"** submitted by **Mr. Lokesh Jain (Roll no. 2K17/PhD/CO/10)** as a Research Scholar, Department of Computer Science, DTU, Delhi is carried out by the candidate under our supervision and guidance.

The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Prof. Rahul Katarya (Supervisor)
Department of Computer Science & Engineering
Delhi Technological University,
Delhi, India-110042

Dr. Shelly Sachdeva (Co-Supervisor)
Associate Prof. & HOD, Department of Computer Science & Engineering
National Institute of Technology,
Delhi, India-110040

ACKNOWLEDGMENT

I address my sincere thanks to Almighty God for giving me the inner power to complete my dissertation and guide me in every step of my life.

It is an immense pleasure to have the opportunity to extend my heartiest gratitude to everybody who helped me throughout this research. It is a distinct pleasure to express my deep sense of gratitude and indebtedness to my learned supervisors Prof. Rahul Katarya (Dept. of Computer Science & Engineering) and Dr. Shelly Sachdeva (Associate Prof. & HOD, Dept. of Computer Science & Engineering, NIT, Delhi) for their invaluable guidance, encouragement, and patient reviews. During the research, their motivation and encouragement made me go further than expected in the development process and the redaction of this dissertation. I am deeply indebted to both of my supervisors for guiding me in carrying out the research work and morally supporting me in every way during the course's challenging times. Their technical expertise, precise suggestions, kind nature, and detailed, timely discussions are wholeheartedly appreciated.

Also, my sincere thank goes to Delhi Technological University for considering my candidature for this course. I am also very thankful to Prof. Yogesh Singh, Vice-Chancellor, Delhi Technological University, Delhi, who has been a constant source of enthusiasm. He has always motivated young researchers like me to pursue excellence to achieve higher goals in academics and research. Also, my sincere thanks reciprocate to Dr. Rajni Jindal (HOD & Chairperson DRC, Dept. of Computer Science and Engineering) and Dr. Ruchika Malhotra (Ph.D. Coordinator) for insightful comments and valuable suggestions. Special thanks to my seniors and colleagues of IITM, Janakpuri, Delhi. My sincere thanks to all the professors, faculty, researchers, and nonteaching staff of the Computer Science Department.

I also wish to take this opportunity to thank all my teachers who have taught me and shaped me into the person I am, aggravated me to be an academician, and have directly indirectly made me capable of succeeding in completing this research work.

Most prominently, my heartfelt gratitude is for my parents, who are the motivations behind me; without their blessings, this work could not have been accomplished. I am deeply grateful to my family for their moral and emotional support in my life and had to miss lots of occasions of being with me during this work.

Lokesh Jain
Department of Computer Science
Delhi Technological University, Delhi
Email: *nsitlokesh.jain@gmail.com*

ABSTRACT

"A leader is one who knows the way, goes the way and shows the way."

— ***John C. Maxwell***

Context: In the current scenario, the online social network has plenty of perspectives to interact with the other person, share information and ideas, and discover and realize the new thing within a single click. Social networking sites provide web-enabled resources that make human life more convenient and contented through the more comfortable communication method. Whenever people find any decisive situation in their daily lives, they share their viewpoints and opinions through blogs, social networking services, reviews, pictures, videos, etc. In the social network, an opinion leader is a person who has the ability to deflect human decision-making through their dexterity, knowledge, experience, and attitude. Nowadays, organizations appoint opinion leaders to promote their products as part of marketing strategies. The applicability of opinion leaders is very abundant in real-world applications like marketing, finance, recommender system, healthcare, consumer behavior, online learning and knowledge communities, blogosphere, and many more diverse fields. Therefore, this study represents an organized, systematic, and arranged effort that determines the identification, power, and applicability of opinion leaders in the online social network.

Objective: The objective of the entire study has been classified into three segments.

- The primary objective of the study is to discover the optimal opinion leaders in the online social network based on computational intelligence techniques.
- The second objective focuses on presenting the significance and power of opinion leaders for the diffusion of products in the online social network.
- The third objective is exploring the applicability of the opinion leader through the online social network in healthcare.

Methodology: For achieving the mentioned objectives, this study utilizes computational intelligence techniques like nature-inspired metaheuristic algorithms, game theory, graph neural networks, and fuzzy logic due to the tremendous applicability to solve natural world problems. Following strategies are used to achieve the targeted objectives:

- For achieving the first objective, innovative and novel computational intelligence techniques are implemented to identify the suitable opinion leader in the online social network. The social network-based variant of two nature-inspired metaheuristic algorithms, the firefly algorithm and whale optimization algorithm, respectively, are used to find the opinion leaders.
- To accomplish the second objective, two approaches; Game theory and Graph Neural Network-based, have addressed the importance and power of the opinion leader for the diffusion of products in the online social network. The game theory-based strategy is used to elucidate the coalition of opinion leaders, while Graph Neural Network-based technique proposed a reputation and trust-based unique model to show the relevance of opinion leaders for information diffusion.
- To attain the third objective, the relevance and applicability of opinion leaders are explored in healthcare. In the ongoing Covid-19 pandemic course, people spread various COVID-19 related rumors and hoaxes on social networks, which incredibly negatively influences civilization. A reputation-based opinion leader identification algorithm is designed that identifies opinion leaders to control the spreading of COVID-19 rumors.

Results: The outcomes of the study are as follows:

- A social network-based firefly algorithm is designed to identify the opinion leader within the local communities and globally. The Accuracy, Precision, Recall, and F1-score of the firefly-based model is 0.94, 0.93, 0.95, and 0.94, respectively.
- A modified Louvain community partitioning algorithm has been designed to identify the communities in the network. The algorithm’s modularity gain and running time are around 14% lesser than the original Louvain method.

- A social network-based whale optimization algorithm is addressed that suitably recognizes the opinion leaders based on various optimization functions. The Accuracy, Precision, Recall, and F1-score of the model are 0.95, 0.94, 0.95, and 0.95, respectively, based on 12 standard benchmark functions.
- An innovative community partitioning algorithm is also designed to find the communities based on neighborhood similarity. The total running time of the algorithm is 11% faster than the other standard algorithm. Also, the other parameters like Node attribute similarity, Common neighbor similarity, Average community density, etc., are reduced to around 12%.
- A new Game theory-based Opinion Leader Detection algorithm is presented to identify the coalitions of opinion leaders with the maximum synergy, marginal payoff, and Shapley value. The Accuracy, Precision, Recall, and F1-score of the model are around 0.94, 0.94, 0.95, and 0.95, respectively. Also, the diffusion rate is enhanced by approximately 15% over other SNA measures.
- An exclusive Graph Neural Network (GNN) for Opinion Leader Identification (GOLI) model is proposed that utilizes the power of GNN to categorize the opinion leaders and their impact on the diffusion of products in the online social network. The GOLI model obtained around 91% training accuracy and 92% testing accuracy with an approximately 1% error rate. The Accuracy, Precision, Recall, and F1-score of the model are around 0.95, 0.96, 0.96, and 0.96, respectively.
- An Opinion Leader-based Rumor Detection (OLRD) algorithm is designed to show opinion leaders' applicability and significance for controlling the COVID-19 rumors. The proposed approach reduced the total number of diffusers by 26% faster, spread veracity around 22% more quickly, and impacted approximately 23% faster than other SNA measures.
- A Reputation-based Opinion Leader Identification (ROLI) algorithm is defined to find the opinion leaders in the online social network. The proposed model produces 91% Accuracy,93% Precision, 95% Recall, and 94% F1-score.

Table of contents

CHAPTER 2: METHODICAL LITERATURE REVIEW

CHAPTER 3: OPINION LEADER IDENTIFICATION USING THE NATURE-INSPIRED META-HEURISTIC OPTIMIZATION ALGORITHMS

CHAPTER 4: RECOGNITION OF OPINION LEADER'S COALITIONS FOR INFORMATION DIFFUSION: GAME THEORY APPROACH

CHAPTER 5: OPINION LEADERS FOR INFORMATION DIFFUSION USING GRAPH NEURAL NETWORK

CHAPTER 6: APPLICABILITY OF OPINION LEADERS TO CONTROL COVID-19 RUMORS IN ONLINE SOCIAL NETWORKS

CHAPTER 7: CONCLUSION AND FUTURE SCOPE

LIST OF ABBREVIATIONS

2-D	2-Dimension
AI	Artificial Intelligence
AIS	Artificial Immune System
ANN	Artificial Neural Network
BC	Betweenness centrality
CC	Closeness centrality
CI	Computational Intelligence
CNN	Convolutional Neural Network
COVID-19	Coronavirus Disease
DC	Degree centrality
DF	Degree of Freedom
DT	Direct trust
EA	Evolutionary algorithms
EC	Eigenvector centrality
FL	Fuzzy Logic
FNN	Feedforward ANN
GA	Genetic Algorithm
GA-ANN	Genetic Algorithm-Artificial Neural Network
GA-SVM	Genetic Algorithm- Support Vector Machine
GNN	Graph Neural Network
GOLI	GNN for Opinion Leader Identification
GPU	Graphical processing unit
HOPE	High-Order Proximity preserved Embedding

IDT	Indirect trust
LLE	Locally Linear Embedding
LSTM	(Long / Short Term Memory
MS	Mean Square
OL	Opinion Leader
PR	Page Rank
RA	Reputation Aggregator
RNN	Recurrent Neural Network
RT	Recommended trust
SNA	Social Network Analysis
SS	Sum of Squares
SVM	Support Vector Machine
WWW	World Wide Web

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In the current web era, social media's role is essential and vital in individual life. Life without social media is like a land without water. Similarly, opinion leaders' presence in human life is critical and significant for information diffusion and the mind-making process. Section 1.1 covers the detailed view of online social networks along with the significance. Also, different types of online social networks also discussed in detail. Section 1.2 describes the opinion leaders' definition, the need for opinion leaders, and opinion leaders' characteristics are explained in detail. Section 1.3 illustrates the overview of computational intelligence techniques. Section 1.4 covers the motivation behind the study. Section 1.5 wraps the research questions and research objectives in detail. The brief of all the chapters is discussed in section 1.6, while in section 1.7, the summary of the entire chapter is concerned.

1.1 Online social network

Utilizing IT Solutions and Web-based applications in individuals' human lives cause the age of the social network's broad scope. The practical investigation and analysis of social networks are generally significant to recognize the interaction and association among the people. Typically, the network organization comprises a bunch of nodes and links demonstrating the different kinds of connections between the nodes.

The primary inspiration of social networks is derived by J. A. Barnes, an anthropologist, who analyzed and investigated human behavior and attitude in intellectual societies and communities in 1954 [1]. For the last many decades, social networks pay attention to the domain of sociology, psychology, geography, graph theory, neuroscience, statistics, producer-consumer behavior science, and computer science .

In 1967, Stanley Milgram proposed the theory of contemporary social networks and explained the concept of small-world networks [2]. According to this theory, each person in the world is indirectly connected, i.e., if there is no straightforward relationship between user i and j, both can share the stable connection through the third user who knows both of them. Although it is a fascinating and innovative concept, the other contributors did not readily accept the idea. During the research, Milgram performed two operations; in the former one, he asked the person to send the letter to another person who lived in another secluded city. One condition associated with the letter transmission is that it should have been delivered only to the person directly connected to the previous person who offers the letter to him. He found that the people in one city created a social network with the person in another city, i.e., a network is formed based on the user's interactions. He also analyzed 'six degrees of separation' between the source and the designation user, i.e., only five intermediate people must deliver the letter from one person to another person [3].

Afterward, Pool and Kochen presented Milgram's theoretical prototype and demonstrated the fundamental research in a graphic pattern through various simulations and experiments [4]. They also explained that if the selection criterion to choose the source is known with some preliminary information, the six degrees of separation would be reduced in some cases. Further, other researchers also investigated that the six degrees of separation are sufficient for offline communication, while for online communication, only three degrees of separation are genuinely appropriate for transmission [5], [6].

With the evolution of internet technology and the World Wide Web (WWW), collaborative computing, also known as social computing, materialized as a surging and potential field of learning in the domain of computer science [7]. Collaborative computing involves the various networking services and tools, cooperative groups and forums, embedded software and hardware, and specialist and experts [8]. In the present time, collaborative computing is also referred to as social networking that includes data accessing, mining, gathering, processing, analyzing, and revealing different types of social data [9].

Thus, the leading concept of the social network is derived from the term 'society.' A society is not just a collection of person or individual; it is slightly the aggregation of views and ideas shared by the people and establishes the connection among them [10]. Theoretically, a social network is delineated as a collection or accumulation of users and their relationship connecting these users [11]–[14]. Various researchers addressed social network definition in slightly different ways. So, both formal and informal descriptions of the social network have been discussed by researchers. According to Wasserman and Faust et al.[15], A social network is a collection of nodes in which each node represents the actor. He is an individual, collaboration, or mutual collective unit, while a relationship is depicted as the linkage between the actors. For example, the companionship among the students in the school, friendship among the employees working in the same team or organization, etc. According to Hanneman and Riddle et al. [16], a node indicates the point or agents and various types of multiple connections established among the issues. For example, the relationship between two agents, for example, the manager and staff member, can be friendship, neighbor, or team member. According to Kukla et al. [17], a social network is an aggregation of people and connections. If the people performed the same event continuously or periodically, they formed the formal or informal community based on frequent interactions. According to Yang et al. [18], an actor in the network is treated as a customer. The connection between the customers is represented by connectedness that shown the strong tie between the customers based on their business benefits. For example, companies alliances and make a network for the growth perspective. Liben-Nowell et al. [19] addressed a new definition of the social network in which people connected according to the specific reason, benefit, and domain of interest. For example, a business person interacted to enhance product productivity, sales and focus on business growth.

1.1.1 Why study online social networks?

Due to the ongoing growth and innovation on the internet and web 2.0, various services and facilities are offered and provided by the computer network. The group of people who used these services formed a virtual network, generally known as an online social network. The online social network is also referred to as a virtual community [20], web community

[21], computer-supported social network [22], or web-based social network [23]. The online social network provides different kinds of services such as online views, images, music, and videos sharing, conferences, meeting, product promotion, online gaming, learning communities, blogs, research collaboration, business, employee-oriented services, social groups, message services, and many more as shown in Fig. 1.1 [24]–[26].

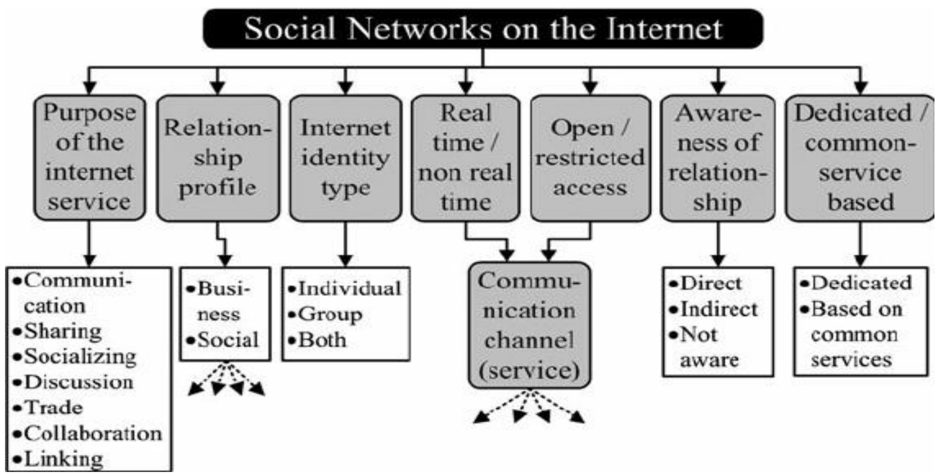


Fig. 1.1. Categorization of Social Networks on the Internet

1.1.2 Types of online social network

Due to the variety of significance and applications, there are varieties of online social networks in the real world. Generally, based on the node and link attributes, there are two types of social networks: homogeneous and heterogeneous social networks. A homogeneous social network is a collection of similar nodes that share the same properties, while a heterogeneous social network is a group of dissimilar types of nodes. Skype, Facebook, etc., are homogenous social networks that contain the same kinds of users and relationships. Simultaneously, flicker, LinkedIn, etc., are the heterogeneous social networks that include different types of entities and connections. Thus, as shown in Fig. 1.2, based on people's interests, choices, and domain, social networks can be broadly categorized in this manner. A homogeneous social network is controlled by a single authority, company, or society using standard rules and practices based on information availability. The users in

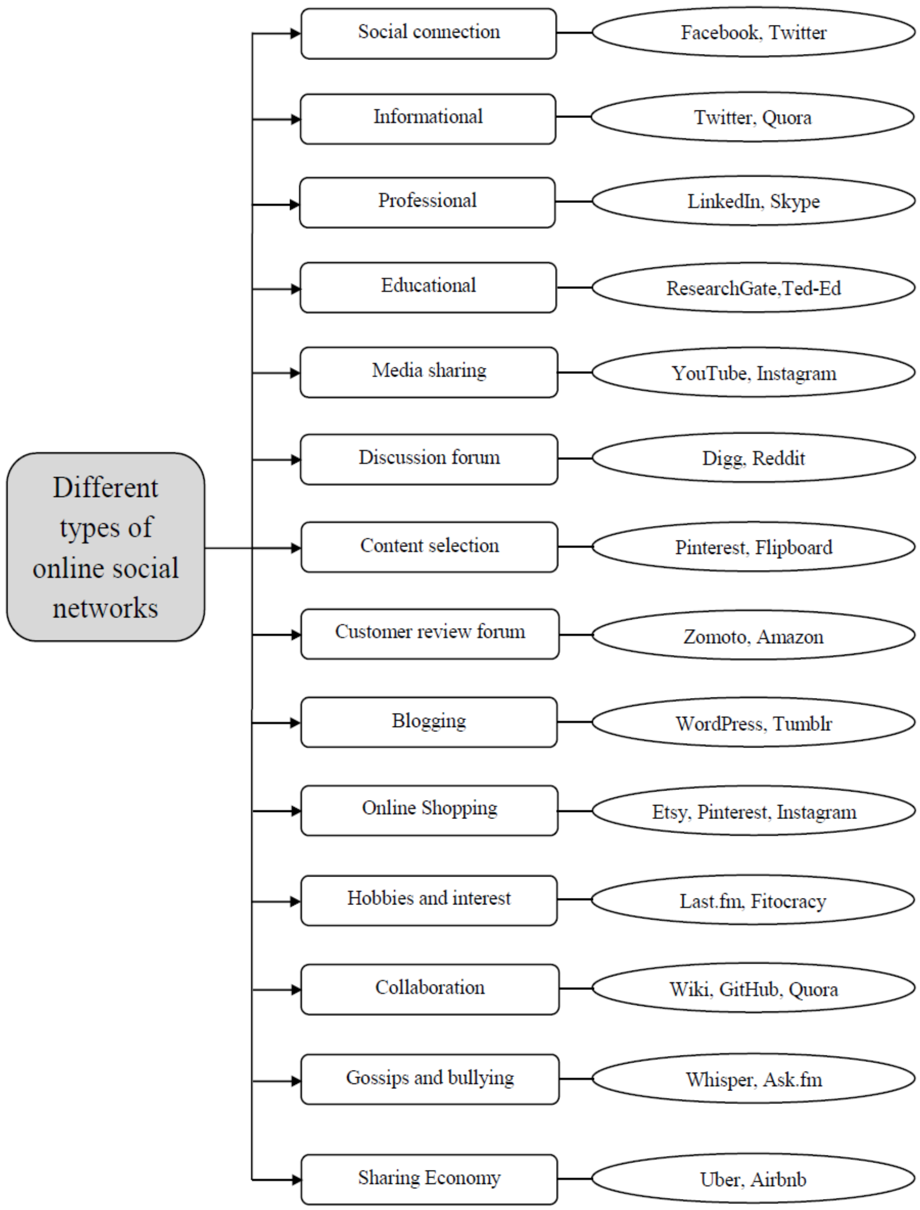


Fig. 1.2. Different types of Social Networks based on interest

the homogeneous social network performed the same activities under the preferred conditions and also shared messages and emails through the familiar interface. On the other side, the heterogeneous social network is considered a multisystem social network.

Multiple entities are belonging to the different layered systems, and all the layers cooperated to make a hybrid system. Numerous connections belong to the multilayer system merge to make a single heterogeneous system that integrated multiple functionalities derived from various platforms [27]–[29].

1.2 Opinion leader

In the 1950s, opinion leaders' character in building and recognizing the individuals' sentiments can be followed back to specific works. Researchers suggested that people are not just chosen for acquiring the prominent administration position dependent on their characteristics; besides, they ought to have a few attributes that coordinate with a specific condition of the community and the bunch of common interests shared by its individuals [30]. Because of their theory, a person or set of the person in the group has the significant capacity to influence the community assessments on some recognizable and applicable issues in contrast with the other person. Katz proposed 'the two-step flow of communication' that significantly addressed opinion leaders' powers for the transition of sentiments and views from social media to the public community [31]. They also considered the influence to integrate three main components; quality of values, capability, and domain of interest. Leader's and follower's character and decisions are likely to change unusually over time and domain applicability [32]. In the era of digital advertising, at whatever point an opinion leader is associated with a specific item or material, these items' buying power expands sooner when contrasted with different things. In this way, an opinion leader impacts another user's conduct by turning over more information about the items through their insight and knowledge [33].

A human being is living in a period that the manners in which individuals speak with one another have been changed drastically due to web-based media destinations' appearance and extension. By proliferation of informal communication locales, assessment sharing sites, online journals, and microblogs, individuals can without much of a stretch and openly collaborate and express their encounters, conclusions, feelings, and sentiments concerning a particular item, administration, or even in a political circle and financial issues. In such a climate that data streams quickly, a few people have a high ability to impact others' assessment or lead them toward a specific point on account of their experience, attitude, seek after objectives, or most likely due to their charming character to incite the feelings of their supporters [34], [35].

In the current scenario, social networks give a fundamental correspondence stage that encourages individuals' associations in the public eye. Because of these connections, others might be influenced by others' suppositions and can further impact them. Among them, the individuals who can move and shape others' assessments are viewed as opinion leaders. Various terminology refers to opinion leaders such as opinion influencers, opinion makers, opinion shapers, opinion former, influence paddler, motivator, etc., based on domain variety and applicability [36].

1.2.1 Types of opinion leader

The development of online communities is increasingly more prominent because of the improvement of the Web. Currently, most users are associated with numerous social networks to share their perspectives and thoughts, such as Facebook, Twitter, Google+, etc. These online social networks give an individual a stage to share their view about any items, images, videos, legislative issues, news, events, and many other related current issues. Sometimes, these online assessment is firmly related with extraordinarily delicate and opposing. Often the information shared on online social networks is not more trusted and verified. Hence, it is obvious to discover trusted and verified thoughts on these gatherings. An opinion leader assumes a significant part to help up correspondence among users and improve the likelihood to spread reliable information.

However, due to opinion leaders' applicability in multiple domains, many researchers considered the opinion leader an influencer [37]. Although influencing other decision-making is one of the attributes of an opinion leader and all the opinion leaders are influencers, every influencer is not an opinion leader. Influence only can persuade other decision-making processes and motivate others to follow their opinions. An influencer can be any family member, celebrity, blogger, YouTubers, sportsperson, etc [38]. Thus, usually, opinion leaders are classified into four categories, as shown in Fig.1.3.

- **Local and global opinion leader**

  The classification of opinion leaders is based on the extent of the impact concerning marketing strategies. The local opinion leaders' scope is limited, i.e., they influenced the people who belong to particular splinter and separated communities with a specialized subject. On the other hand, the power of global opinion leaders impacted a massive amount of followers. They influenced more dense and crowded communities on an international and global scale as per the network structure [39].

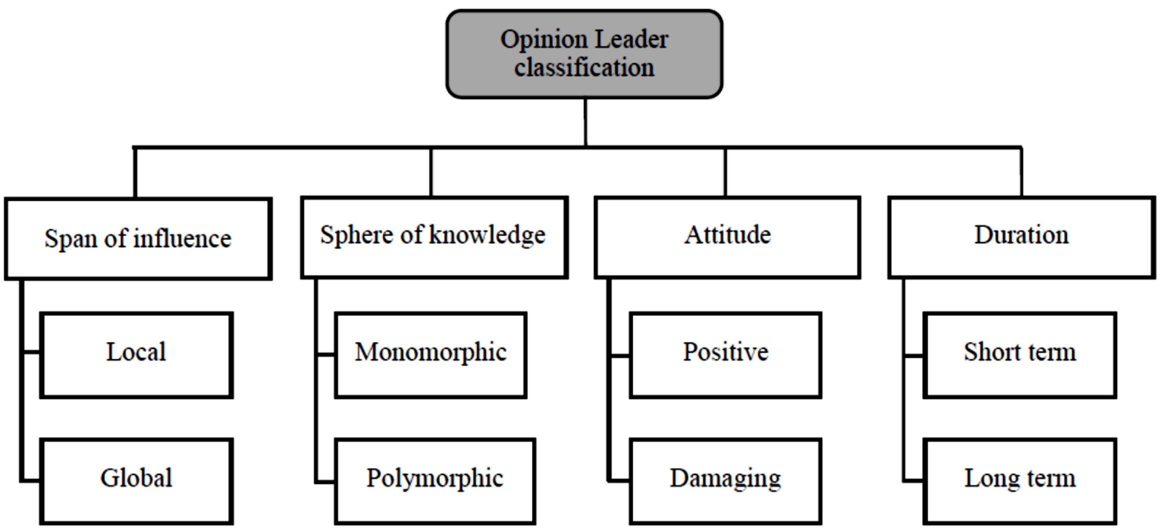


Fig. 1.3. Classification of opinion leaders

- **Monomorphic and Polymorphic opinion leader**

  This classification type is derived from the domain knowledge and specialization from the business and marketing point of view. The discipline and scope of monomorphic opinion leaders are minimal and only on the specific topic. They have profound and insightful knowledge and penetration on the particular subject specifically recommended for a specific society. Polymorphic opinion leaders can spread the information up to a large extent as they have a massive amount of knowledge and information on various domains. According to the analyst, nowadays, most opinion leaders are likely to be polymorphic because of technological advancement and social media sources [40], [41].

- **Positive and damaging opinion leader**

  In the real world societies, often every person has their thoughts and opinion towards particular thing based on internal attitude and external surroundings. Opinion leaders also have a positive and destructive attitude towards the products, issues, and events. Usually, an opinion leader has a positive opinion, but sometimes they also show a harmful or destructive view towards the things they are agreed with or not satisfied with them. The power of damaging opinion leaders is similar to positive opinion leaders. Positive opinion leaders are persuasive and optimistic for promoting the products. Simultaneously, the political parties or the companies adopt damaging opinion leaders to degrade rival companies' exposure or ruling political parties. The destructive opinion leaders are more selfish, gloomy, demanding, compulsive, and self-centered than positive opinion leaders who are very practical, helpful, productive, and goal-oriented [42].

- **Short-term and long-term opinion leader**

  Based on the impact duration, an opinion leader can be broadly classified into two classes: short-term and long-term opinion leader. Suppose the time interval of the opinion leader's influence is around 2 to 3 years, which is needed to form the follower's judgment and decision. In that case, it is considered a long-term opinion leader. On the other hand, if the opinion leaders can quickly penetrate others' decision-making process by experience and skills, opinion leaders are considered a short-term opinion leader. The impact of long-term opinion leaders is long-lasting, deep-rooted, and more impactful than short-term opinion leaders. Besides, the entire manipulation process depends on the user's trust, values, and opinion leader's reputation [43].

1.2.2 Applicability of opinion leader

Since opinion leaders' applicability is very extensive in multiple domains from general product dissemination to critical review and feedback, as per Fig. 1.4, a brief description of the opinion leader's relevant environment is as follows:

- **Digital Marketing:** Digital marketing involves using internet-based utilities, electronic applications, and social media to promote particular products. Opinion leaders can support identifying the most suitable techniques and procedures to spread the product information into large communities [44].
- **Social Media:** An opinion leader is supposed to be a social media model that motivates and propagates almost accurate media information. So, consider the opinion leader is associated with any social media. It is beneficial for the companies to promote their product easily through brand marketing and gain more customers' attention [45].
- **Consumer behavior:** Due to a broader range of knowledge and proficiency, opinion leaders are competent to capture potential customers who are willing to purchase a particular product. In this sense, opinion leaders support both customers and companies to make the correct decision based on the current market trends [46].
- **Online learning community:** Online learning communities provide education facility for users from peer-to-peer learning using the internet and other online platforms. Opinion leaders share their views about the quality and value of the content shared by many users in the form of text, audio, images, and videos. Opinion leaders can also deliver and guide the users to find reliable sources of study and knowledge material significant to a specific group of interested users to gain information about that topic [47].
- **Agriculture:** As per the recent trends, the opinion leader acts as a mentor who has a significant social and economic growth role. Opinion leaders have strong connections with the change agents, innovators, farmers, and social workers. From the economic viewpoint, the opinion leaders once adopt the innovation multiply the whole efforts by maintaining the connections with the local workers. Due to the lack of information and technology-enabled resources, people who belong to villages and rural areas are not aware of the new technology-driven transformations and innovations. Opinion leaders

support farmers and land workers by providing valuable agriculture-related information that eventually enhances the national economy [48].

- **Microblogging:** Microblogs provide a platform to post and share a small quantity of relevant and current information among the audience quickly in the form of text, audio, video, and pictures. The opinion leader uses the microblog for information dissemination and identifying the current trending events and news. Through microblogging, the opinion leaders are evaluated and analyzed the customer feedback and reviews of the material, consumer preferences, and choices profoundly that help companies upgrade the features and quality of products through innovative ideas [49].
- **Online Knowledge community:** Online knowledge communities build up a society where various researchers and scholars belong to different regions on an online assembly platform and share knowledge on the suitable domain. An opinion leader is a primary user who controls and transfers the appropriate information intelligently in the other communities. So, it is rather significant to transmit only relevant information that is effective to a particular group of users [50].
- **Online review community:** Online review communities are essential to establish a connection, trust, fidelity, and force sales of the organizations. To enhance business growth, the opinion leader initiates the strategies that generate the user-centric content, monitor user behavior, robust reputation system, embed with the existing system, and perform analysis and customization by user's responses [51] .
- **Recommender system:** due to the massive applicability of the internet, most customers share their views and provide product feedback. But it is challenging to find which feedback is more appropriate and trustworthy. So, the opinion leader supports and recommends customers choose the accurate and correct product so that they do not face any problem due to the quality of the product. On the other side, the opinion leader also helps resolve the cold start problem in the recommended system [52].
- **Blogosphere:** A Blogosphere integrates different types of blogs in which each user shares their views, ideas, and outlook on the specific topic. The aggregation makes a community with varieties of users but interlinked with some common issues. An

opinion leader supports integration by maintaining collaboration among different users to share knowledge and opinion [53].

- **Online social networking:** opinion leaders play a strong character in choosing and sharing the correct information in the social network. If the social networking site often spread misinformation and rumors regularly, it would reduce users' trust in that social networking site, and viewers gradually stop their services [54], [55].

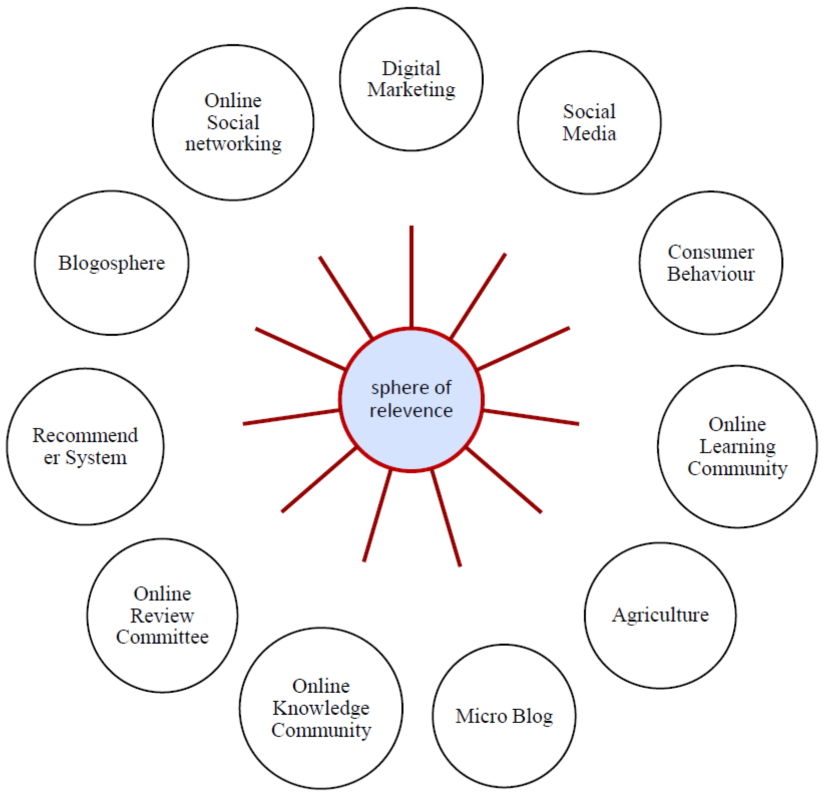


Fig. 1.4. Sphere of the relevance of opinion leaders

1.2.3 Characteristics of opinion leader

Table 1.1 represents the main characteristics of opinion leaders in brief.

Table 1.1. Characteristics of opinion leaders

Characteristics	Description

Exclusive and higher social status	Incredibly decent and estimable because of their specialized information about the item.
Observability	Analyze, detect and examine the issues with a profound sense.
Gregarious	Tend to rise in open groups or unadulterated affiliations due to generosity.
Dogmatism	The propensity to set down standards as undeniably evident, without thought of proof or the other's assessments
Innovative	Lots of original and profitable techniques to promote the product.
Product familiarity	Having a massive amount of information and experience to deal with the product
Risk Preference	They are interested in taking a potential action having high risk but would produce more profitable outcomes.
Leadership	They can control, direct, motivate, and achieve the common objective.
Exposure	They can influence others through rational or experience about any subject related to the domain of interest.

In the Table 1.1, various characteristics of opinion leaders are summarized based on [38], [56]–[58].

1.2.4 The need for an opinion leader

A human is a social creature that lives in society and can share opinions and views on a particular subject. Whenever people face any problem and do not found any solution to solve it, at that time, they are looking for a person or expert who can provide guidance and suggestions to overcome the problem. Consider a case in which a person wants to purchase a mobile phone from an electronic shop. At that moment, you want a bit of advice from an important person on which you trust and provide an evocative suggestion about the car. Thus, at that time, the direction of that person works very well and influences you to make the right decision. So, the opinion leader is an individual who has the quality to guide, assist, influence and provide the solution to the followers and other people. As per the research, opinion leaders have unique characteristics and are useful in various domains. Nowadays, companies adopt opinion leaders to promote products and achieve user trust and credibility [59]. The kind of impact enforced by the opinion leaders on their followers is very exquisite and perfect that exists over a long time. So the power of opinion leaders is very significant to our daily life [60], [61].

In the current time, companies and organizations utilize opinion leaders for designing and shaping the conceptual business model. The abstract business model provides the plan and growth for small and medium businesses and recommends more values and benefits to potential customers. Opinion leaders are involved in the various phases of the model and continuously improve the strategies that lead to pioneering and severe product quality changes [62]. Suppose the businessperson and marketer know their potential customer who follows a particular opinion leader on social media. In that case, they can adopt or contact the opinion leader to influence and advise them to choose the product manufactured by their company [38].

Similarly, in political science, opinion leaders are the key people or associations with an unmistakable status in their networks, which unite, change and engender a particular political pattern to the intended populace in the public eye to advance social agreement and soundness [63]. In healthcare, opinion leaders can help run a proof-based practice; accelerate the dispersion of well-being advancing and sickness forestalling developments [64]. There are many other domains like education, agriculture, marketing, defense, retailing, consumer behavior, and many more where opinion leaders' applicability and need are very significant and vital [65]–[67]. So, in a nutshell, the need for an opinion leader can be summarized as follow:

- Opinion leaders can govern and control other decision-making processes through their skills.
- Marketers utilize the power of opinion leaders to establish and maintain trust and reputation.
- Opinion leaders support the innovative and unique proposal and promote them through Social media services.

- Different types of communities have varieties of opinion leaders, so it is essential to choose the appropriate opinion leader to promote the product effectively.
- Opinion leaders apply advanced technology-driven tools based to diffuse information through social media.
- Opinion leaders have high experience and intelligence to deduce the essence for the followers.
- Opinion leaders are responsible and dedicated to controlling misinformation and rumors on social media.

1.3 Computational intelligence techniques

Logical critical thinking comes from the procurement of information from a particular ambiance, the control of such information, and the involvement in reality with the controlled data. The more thorough and better organized the information base, the more it imitates a logical expansion. The more straightforward arrangement is to investigate more logical issues with sufficient understandings. As the experiences demonstrate, computational intelligence is not just about a machine that works based on predefined instruction. It is about understanding the behavior of agile thought and activity utilizing computer machines as exploratory components [68]. Intelligent devices have been used to arrange better and translate process conduct dependent on the accessible data to acquire input-yield planning and dynamic. The usage of technical information, capacity to utilize loose, dubious data, joining data over different controls, mechanized Artificial Intelligence propelled from the natural world such as neuroscience, and behavioral science improves models for streamlining the framework execution fulfilling the inalienable framework/measure imperatives [69].

Computational Intelligence (CI) integrates intelligent tools, techniques, and mathematical models that accept the unrefined data from the various resources [70]. After distributed processing, it produces outcomes that are incredibly fault-tolerance. CI uses soft computing techniques and approaches such as fuzzy logic, artificial neural network, evolutionary computing, nature-inspired algorithms, swarm intelligence, learning theory, probabilistic mathematical models, and artificial immune system [71] to solve the problems, as shown in Fig. 1.5.

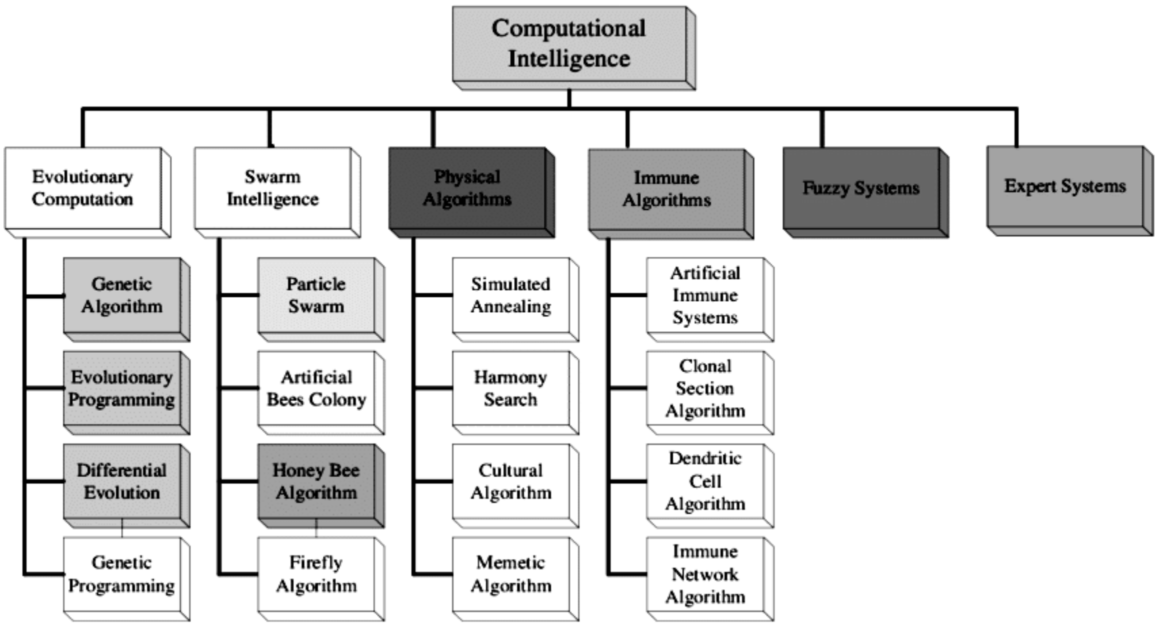


Fig. 1.5. Types of Computational Intelligence Techniques

CI derived intelligence from computer instructions. Bezdek presented the leading phenomenon of CI in 1994: a framework is called computationally smart if it manages low-level information, for example, mathematical knowledge, has an example acknowledgment segment and doesn't utilize data in the AI sense, and when it starts to show computational adaptively, adaptation to internal failure, speed moving toward human-like turnaround and error rates that estimated human routine [72]. Another definition of CI proposed by Bezdek stated that CI is the subset component of AI. AI utilizes hard computing methods to solve the problems, while CI employs soft computing processes to tackle them. Furthermore, AI and CI both have the common long-term objective to solve the problems with intelligence and make a machine perfect to perform any intellectual task that a human being can do with their intelligence [73].

The leading applications of CI include computer science, image processing, power system, healthcare, speech recognition, text mining, video segmentation, Manufacture system, business management, sensor network, engineering management, social science, education, and various other multidisciplinary domains [71], [74]–[78].

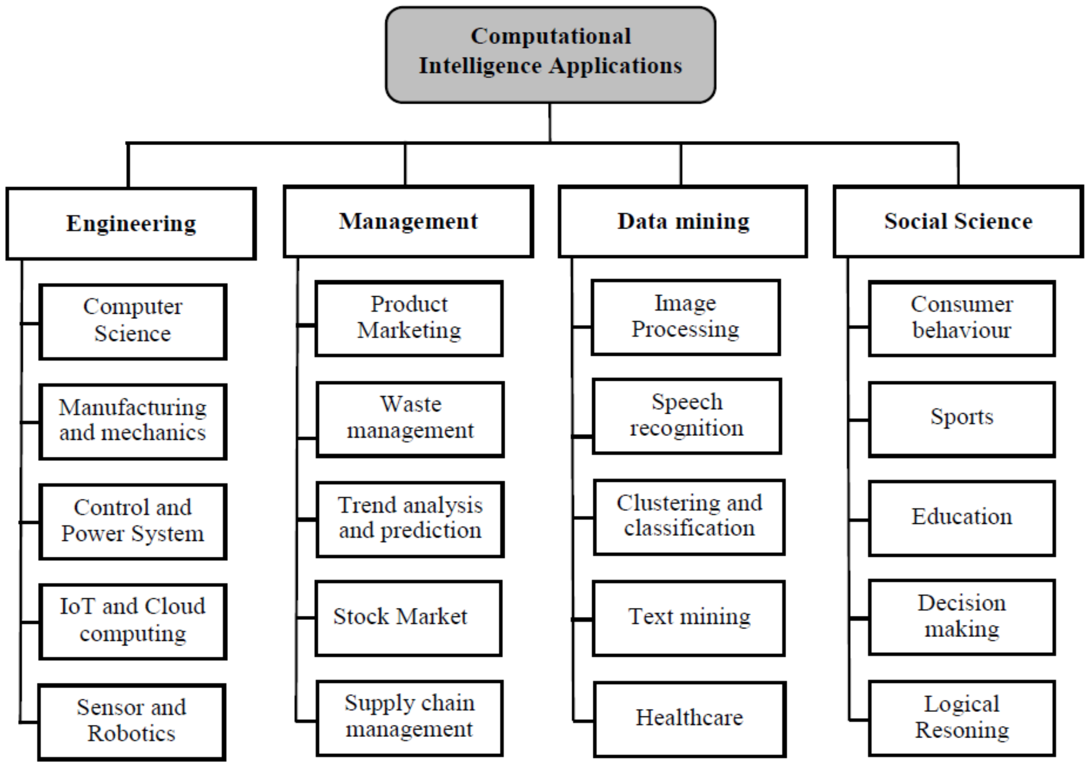


Fig. 1.6. Applications of Computational Intelligence

The most recent CI application includes fault-tolerance machine monitoring using ANN, GA-ANN, SVM, GA-SVM, and other models. ANN-based element removal, speech recognition, noise extraction, power, and control system, sensor and robotics, face recognition, nonlinear optimization, operational task selection and planning, logic processing, granular computing, neuro-fuzzy design, global optimization, behavior analysis, intrusion identification, image segmentation, and many more. Thus, CI emerges as a new discipline of science to solve various multidisciplinary applications. Thus, CI produces better solutions over the conventional hard computing techniques and continuously improves the machine's rational behavior. Therefore the brief description of the main methods and techniques of CI is as follows:

- **Fuzzy logic:** Fuzzy Logic (FL) is a strategy for thinking that looks like human opinion with an unclear and rough statement. The methodology of FL impersonates the dynamic method in people that includes all the intermediary prospects between computerized digital assessment yes and no. The concept of FL is proposed by L. Zadeh, who stated that the traditional computer system only determines the outcomes as yes and no [79]. Still, human reasoning takes the decision consists of all the possibilities between the values yes and no. According to Zadeh, a fuzzy set allows an element to have a partial degree, i.e., the degree of a factor in the group is not entirely right or completely false. If an element x go to a set S, it can be represented as $\mu_S(i) \in [0,1]$ the membership of the element i in set S can be represented as A = (i, $\mu_S(i) \mid i \in X$) Where X represents the universe of discourse .
- **Artificial Neural Network:** An Artificial Neural Network (ANN) is an organic stimulated computational network analyzed like the human brain and executes the information accordingly [80]. It is the pillar of AI to solve various real-world problems that are possibly unfeasible by traditional models and techniques. ANN consists of different neurons known as processing units that communicate and feed information to process the input. Every processing unit is connected with other units through links, and weight is associated with the link for learning purposes. There are hundreds or thousands of hidden layers between the input and output layers. The output of each layer is forwarded as input for the next layer. There are a variety of ANN like CNN (Convolutional Neural Network), RNN (Recurrent Neural Network), FNN (Feedforward ANN), LSTM (Long / Short Term Memory), and many more based on the different type of applications. The application of ANN includes Aerospace, defense, finance, stock marketing, medical, industries, speech, telecommunication, signal processing, anomaly detection, control system, voice recognition, text analysis, and many more [80], [81].

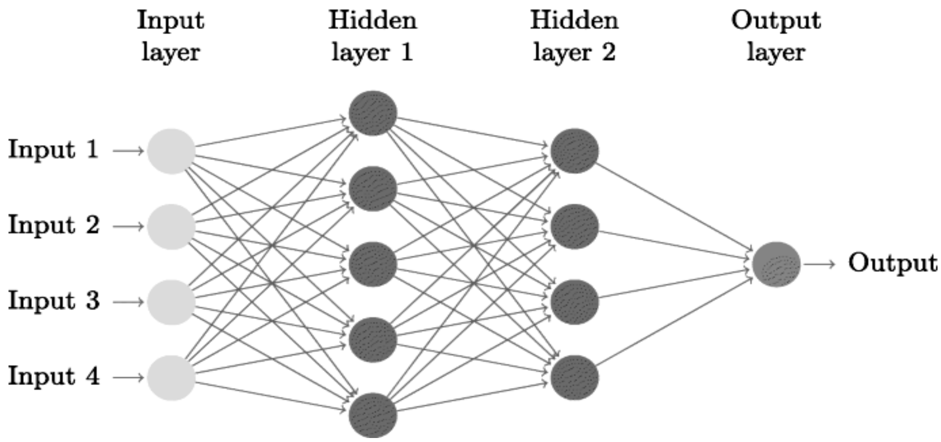


Fig. 1.7. Structure of ANN

- **Evolutionary Algorithm:** Evolutionary algorithms (EA) are heuristic-based techniques to deal with taking care of issues that can not be completed in polynomial time, for example, traditionally NP-Hard matters, and whatever else that would take dreadfully long to measure comprehensively [82]. They are regularly applied to combinatorial issues; in any case, hereditary calculations are frequently utilized pair with different strategies, going about as a quick method to discover a, to some degree, the perfect beginning spot for another estimate. In EA, there are a total of five necessary steps: Initialization, selection, Mutation, crossover, and termination. Initially, find the problem's population, see the possible solutions, and calculate each possible solution's fitness to verify how perfectly they perform. In each iteration, the population changes with time and produces a new potential solution. Next, the possible solution is chosen based on the highest fitness function score and mutation. In the crossover operation, two or more than two base solutions are selected, and perform the off-springs operation to generate the new solution using the parents' content. In mutation, a variant n-ary operator operates on the chosen solution and create a new aspirant resolution. Finally, in termination phase, all the remaining solution terminated, and only those solution exists having the maximum fitness value. The same process is iterated multiple times until the best solution is found.

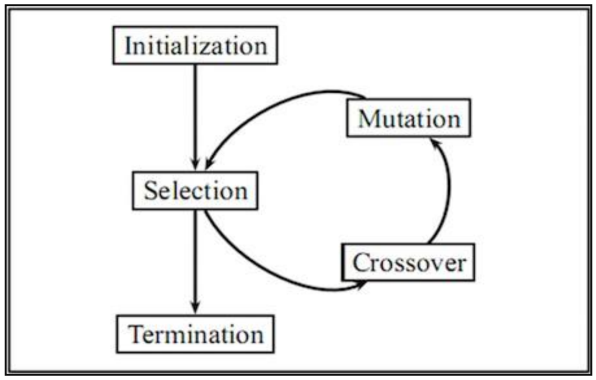


Fig. 1.8. Phases of Evolutionary Algorithm

- **Learning theory:** Learning theory is the discipline of AI that includes the mathematical, statistical, and conceptual model for enumerating the learning problem. Computational learning theory is a novel and fast intensifying investigation domain that analyzed recognized models of stimulation by way of the objectives of determining the standard techniques underlying well-organized knowledge procedures and discovering the computational obstructions to learning.
- **Artificial Immune System (AIS):** AIS is a branch of CI that is considered as the computational system that is originated by hypothetical immunology and monitored experiments, principles, and frameworks used to solve complex real-world problems. AIS is an adaptive, self-regulatory, distributed, resistant, responsive, and extractor that supports and simplifies the biologically related problem. The main application of AIS pattern matching, feature mining, information mining, assortment, distributed processing, self-protection, dependable system, self-healing system, etc. [83].

1.4 Motivation of study

Both opinion leaders and opinion receivers/searchers have their purposes behind giving data and accepting/looking for item data and guidance. Opinion leaders offer item-related data and advice now and again deliberately and some of the time when they are drawn closer and requested. Practically, opinion receivers/seekers demand data or tune in with persistence to all that the opinion leader needs to state. There are different reasons why such correspondence trade happens between opinion leaders and opinion receivers/seekers, family members, companions, associates, or even outsiders. This way, they embraced the items to shoppers and, by implication, advance and profited the organization through the online social network. They are also backing to dodge any rumor by distinguishing the data's integrity that may put humans in danger. These clarify the intentions behind the opinion leader recognition. Thus, in a nutshell, the motivation behind this research is summarized as:

- **6-R Concept:** The phenomenon of 6-R includes Right purchase decision, Right product and service, Right brand, Right price, Right store, and Right time. The basic meaning of 6-R is that usually, most people do not have the right decision to purchase the right product and use the right brand's right services at the right cost and time. So there is a need for a study that presents a unique solution to find the opinion leaders in the online social networks.
- **Penetration of E-Commerce:** Due to the extensive visibility and promotions of multiple E-commerce websites, it is very tough to find which website provides the best products to the customers at the right cost. Thus, there is a need for a person who can support choosing the right E-commerce services so that the customers can purchase the products effectively.
- **Lack of knowledge:** Most of the customers do not have much information about marketing strategies and policies. They are easily influenced by any of the unreliable and untrustworthy sources. So there is a need for a trusted and authenticated entity that can spread only the reliable and trustworthy information generated from the authenticated sources after validating its veracity. The customer may not fall into any deep trap.
- **Cold start problem:** The cold start problem is related to the recommender system that stated that if a new product launches in the market and only a limited amount of users have the information about the product quality. It is challenging to promote the product due to a shortage of user ratings and recommendations. Although the new development

has more features and grades than the same existing products, it is far from its actual growth due to a lack of marketing and promotions. Thus, it is required to find a person who can endorse and promote the other people with their skills and experience.

- **Marketing strategies:** In the current competitive environment, if a person wants to initiate the startup or introduce a new product in the market, it very critical to establish credibility and loyalty with the customers due to adapting and time-varying strategies change over time. Thus, finding the right time to follow a correct approach is very critical and time-consuming. Opinion leaders can help and support organizations to think out-of-box and find out the solution to easily promote and deliver the product information visible to all the potential customers worldwide through social media tools and services.
- **The requirement of innovative techniques:** As the technology and generation evolve, various novel and innovative designs, models, and approaches such as deep learning, evolutionary computing, optimization techniques, soft computing, NLP, sentiment analysis, virtual agents, marketing automation, social network analysis, etc., smoothly working with the current system. Various other modules dependent on these innovations developed to solve multiple complex systems significantly. Hence, there is a need for various novel models, algorithms, and approaches based on the new technology-driven platform to detect the opinion leaders in the online social networks. Besides, these methods would also demonstrate the applicability of opinion leaders in various domains.

1.5 Research objectives

This research’s leading objective is to try to overcome the limitations and design novel and innovative methods that inspect the technology-driven techniques. Thus, in order to achieve the designated aim, the following Research Questions(RQs) are identified:

- RQ1: What should be the methods to find out the opinion leaders in the online social network based on computational intelligence techniques?

- RQ2: How can we improve the accuracy of the methods?
- RQ3: What is the role and power of opinion leaders in information diffusion?
- RQ4: What are the techniques to aggregate the collaboration of opinion leaders?
- RQ5: How to explore the applicability of opinion leaders in healthcare through the online social network?

for finding the solution to the above sub-queries, the following Research Objectives(ROs) are finalized:

- **Research Objective 1:** Implement suitable computational intelligence techniques to identify the opinion leader in the online social network.
- **Research Objective 2:** Addressing the significance and power of the opinion leader for the diffusion of products in the online social network.
- **Research Objective 3:** Exploring the applicability of the opinion leader through the online social network in healthcare.

The detailed description of the identified Research Objectives are as follow:

- **Research Objective 1:** In this objective, innovative and novel computational intelligence techniques are implemented to identify the online social network's suitable opinion leader. Due to the versatility of CI to resolve many real-world problems, two nature-inspired metaheuristic algorithms have been discussed to find the opinion leaders. The leading concept of nature-inspired algorithms is generated from the behavior of nature's creatures, i.e., birds, animals, insects, etc. Thus, novel and unique meta-heuristic algorithms have addressed to select the opinion leaders from the variety of social networks
- **Research Objective 2:** An opinion leader's function is very central in business marketing and information diffusion. The information diffusion process is helpful to solve very business-related problems. Thus, Game theory and Graph Neural Network-based approaches have addressed the importance and power of the opinion leader to the diffusion of products in the online social network.
- **Research Objective 3:** The relevance of opinion leaders is far-reaching and covers a variety of applications. In this objective, the relevance and applicability of opinion

leaders are explored in healthcare. In the ongoing COVID-19 pandemic course, people spread various COVID-19 related rumors and hoaxes, which incredibly negatively influences civilization. So, there is a need to invent a trust and reputation-based algorithm that identifies opinion leaders to control the spreading of COVID-19 rumors.

In this context, for fulfilling the requirement of RQs and ROs, Table 1.2 demonstrates the mapping among RQs, ROs, and publications.

Table 1.2. Aligning of Research Questions, Research Objectives, and Publications

ROs	RQs	Publication(s)
RO1	**RQ1, RQ2**	• Jain, L., & Katarya, R. (2019). Discover opinion leader in online social network using firefly algorithm. Expert Systems with Applications, 122, 1-15. **[Published, SCIE, IF:5.452]** • Jain, L., Katarya, R., & Sachdeva, S. (2020). Opinion leader detection using whale optimization algorithm in online social network. Expert Systems with Applications, 142, 113016. **[Published, SCIE, IF:5.452]** • Jain, L., & Katarya, R. (2018, February). A Systematic Survey of Opinion Leader in Online Social Network. In 2018 International Conference on Soft-computing and Network Security (ICSNS) (pp. 1-5). **IEEE**. • Jain, L., Katarya, R., & Sachdeva, S. (2019, November). Opinion Leader discovery based on text analysis in Online Social Network. In 2019 4th International Conference on Information Systems and Computer Networks (ISCON) (pp. 446-450). **IEEE.**

RO2	RQ3, RQ4	• Jain, L., Katarya, R., & Sachdeva, S. (2020). Recognition of opinion leaders coalitions in online social network using game theory. Knowledge-Based Systems, 203, 106158. **[Published, SCI, IF:5.921]** • Jain, L., Katarya, R., & Sachdeva, S. (2020). Opinion leader detection using whale optimization algorithm in online social network. Expert Systems with Applications, 142, 113016. **[Published, SCIE, IF:5.452]** • Jain, L., Katarya, R., & Sachdeva, S. Opinion Leaders for information diffusion using Graph Neural Network in Online Social Network. **[Communicated]** • Jain, L., Katarya, R., & Sachdeva, S. (2019, August). Role of Opinion Leader for the diffusion of products using Epidemic model in Online Social Network. In 2019 Twelfth International Conference on Contemporary Computing (IC3) (pp. 1-6). **IEEE.**
RO3	RQ5	• Jain, L., Katarya, R., & Sachdeva, S. (2020). Recognition of opinion leaders coalitions in online social network using game theory. Knowledge-Based Systems, 203, 106158. **[Published, SCI, IF:5.921]** • Jain, L., Katarya, R., & Sachdeva, S. Impact of Opinion Leader to control Covid-19 rumor in Online Social Network. **[Communicated]** • Jain, L., & Katarya, R. (2018, December). Identification of opinion leader in online social network using fuzzy trust system. In 2018 IEEE 8th International Advance Computing Conference (IACC) (pp. 233-239). **IEEE.** • Jain, L., Katarya, R., & Sachdeva, S. (2019, August). Role of Opinion Leader for the diffusion of products using Epidemic model in Online Social Network. In 2019 Twelfth International Conference on Contemporary Computing (IC3) (pp. 1-6). **IEEE.**

		• Jain, L., Katarya, R., & Sachdeva, S. (2021, April). Applicability of the Opinion Leader to spread COVID-19 vaccine awareness through Online Social Network based on Sentiment Analysis. In 5th International Conference on Advances in Computing and Data Sciences (ICACDS). **Springer [Accepted]**

1.6 Outline of the thesis

The thesis consists of six chapters describing the entire study in a very concise and precise way. A brief description of each chapter is as follows:

- **Chapter 2:** In this chapter, all the state-of-the-art methods and previously developed methods and their merits and demerits to find the opinion leaders in the online social networks have discussed. During the research, multiple approaches are studied to find the opinion leaders in the OSN; Thus, the chapter also covers the background details of methods and Social Network Analysis (SNA)measures to find the opinion leaders.
- **Chapter 3:** In this chapter, two nature-inspired meta-heuristic social network-based approaches, firefly, and whale optimization, respectively, discuss that optimally find the list of top-n opinion leaders in the different online social networks. The time complexity and advantages of both approaches are explained in detail. The proposed techniques are implanted on different datasets. For validating the authenticity of the methods, the experimental outcome is compared with the other standard SNA measures that illustrate the supremacy of both the approaches over other actions.
- **Chapter 4:** In this chapter, a game theory-based approach is addressed that defines the coalitions of opinion leaders based on synergy and Shapley value. The game theory based practice explores the mathematical and logical models that companies use for the decisive cooperative system. Four alternative solutions have been provided to find out each user's payoff in the clusters of the different kinds of online social networks.

Experimental results also prove the power of opinion leaders for the diffusion of products and information.

- **Chapter 5:** In this chapter, a Graph Neural Network-based model has been proposed that pulls up a new horizon to find the opinion leaders and show their powers for the diffusion of new products. A novel method is addressed to calculate the node's trust and reputation differently from the traditional network embeddings approach. The experimental results illustrate the importance of the proposed model over other SNA measures.
- **Chapter 6:** The importance of opinion leaders is vast, deep, and multi-directional, So, in this chapter, the significance of opinion leaders is explored in healthcare to control the COVID-19 rumors to avoid the risk of mental and physical health damage. A novel reputation-based Vote-Rank algorithm is designed to find the prominent opinion leaders who verify tweet's veracity through entropy calculation and sentiment analysis. Twitter's COVID-19 tweets dataset is used for the experimental purpose, and the performance of the proposed approach is measure through various performance metrics.
- **Chapter 7:** In this chapter, the current research's future scope and limitations have been discussed. In the present time, deep learning and graph embedding based models produce better results on the social network datasets. Finally, a few futuristic suggestions and plans have been suggested to find the opinion leader more precisely as the future of SNA is very bright and strongly opens new flaps for the companies.

1.7 Chapter summary

This chapter covers the overview of the entire work along with the content and description of each chapter. Each chapter includes some unique concepts and ideas to support the title and objective of the thesis strongly. The brief of the online social network, opinion leaders, and computational intelligence is also explained in the chapter. The main objectives and scope of the entire work with motivation to study are also described in detail.

Chapter 2

Methodical literature review

With the advancement of technologies, many approaches have proposed identifying the opinion leader in the online social network. In this chapter, different methods are analyzed and investigated based on various parameters such as centrality, power, trust, responsiveness, social status, retweets, topic sensitivity, posted text content, responsiveness meta-heuristic algorithms, and many more to classify the opinion leaders. Section 2.1 covers the overviews of the chapter. Section 2.2 covers the Social Network Analysis (SNA) measures background includes different centralities and metrics. Section 2.3 explores the entire review process, i.e., study plan, selection of studies, and dataset extraction. Section 2.4 explains the considerable description of the studies and their merits and demerits in opinion leader identification. Section 2.5 describes the research gap and limitations of the previous studies. Section 2.6 concludes with the summary of the chapter.

2.1 Overview

The technological movement and revolution play a crucial position to identify the opinion leaders in the social network. An opinion leader can be detected qualitatively and quantitatively based on the domain of interest need. The conditions for choosing the opinion leaders may vary in each method. The social network studies follow meticulous approaches to identify the actor or person who catches the public attention, having more reputation, and having a higher degree of connections with other actors. These actors have higher chances to become the center of attraction and gain affluent public opinions in the network. The social network theory follows the 'centrality' concept to identify the total number of connections and degrees with other entities. Formally, these centralities measures are known as SNA measures that include various kinds of calculative measures.

For quantification, the most prominent nodes in the network's centrality analysis must achieve the desired goal.

2.2 Social network analysis

The Social network can be well-defined as per the undirected graph G = (V, E), where V signifies the group of nodes indicating actors or user and E denotes the group of edges showing the relationship between the actors. The relationship between the actors can be friendship, acquaintances, author-coauthor relationship, and many more. SNA is the system of investigating social structures through the usage of networks and graph ideas. It characterizes networked systems in phrases of nodes (person actors, people, or things within the network) and the ties, edges, or hyperlinks (relationships or interactions) that join them. The leading focus of SNA is to identify the dynamics and characteristics of the network. In SNA, a node's centrality component is suitable to classify the most convenient and influential node in the network [84]–[86]. Furthermore, the four types of different centrality measures are used for calculating the reputation of the node.

2.2.1 Centrality

The power of centrality is imperative and considerable for measuring the significance of node and edge in the network. Four types of different centrality measures, such as closeness centrality (CC), betweenness centrality (BC), eigenvector centrality (EC), Degree centrality(DC), and PageRank (PR) [87], is used for calculating the prominence of the node in the network.

2.2.1.1 Closeness centrality

Closeness centrality (CC) grooves a value for each node based on their 'closeness' to all other nodes within the network. This scheme computes the shortest paths between all nodes and, based on the sum of the shortest path, assigns a value to every node, as shown in Eq.(2.1).

$$CC(x) = \frac{1}{\sum_{y=1}^{y=n} d_{(x,y)}} \qquad \ldots(2.1)$$

Where $d_{(x,y)}$ is the shortest distance between node x and node y.

2.2.1.2 Betweenness centrality

Betweenness centrality (BC) considers the degree that counts the occurrence of a node on the straight path between other nodes. It can be defined as the fraction between the total number of shortest with existing between node i and j passes through node x to the total number of the shortest path between node i and node j as shown in Eq.(2.2).

$$BC(x) = \sum_{i \neq x \neq j} \frac{c_{(i,j)}(x)}{c_{(i,j)}} \qquad \ldots(2.2)$$

Where $c_{(i,j)}(x)$ represent the total number of the straight path between node i and node j.

2.2.1.3 Eigenvector centrality

Eigenvector centrality (EC) is used to quantify the impact of a node in the network. In this scheme, a relative mark is assigned to all nodes, constructed on the knowledge that Association to high-marking nodes contributes more to the score of the node than equal connections to low-marking nodes. EC shows that an actor is all the more vital as it is associated with actors who are themselves essential, as shown in Eq.(2.3).

$$EC(i) = \frac{1}{\lambda} \sum_{j \neq i} y_{ij} . x_j \qquad \ldots(2.3)$$

Where y_{ij} is the adjacency matrix and node x_j is the neighbor of node x_i.

2.2.1.4 PageRank

PageRank centrality (PR) is used to uncover influential or important nodes whose influence extends beyond just their straight acquaintances, as shown in Eq.(2.4). This algorithm is used in the citation network, monitoring network activity, etc.

$$PR(x) = \alpha + \sum_j a_{ij} \frac{x_j}{L(j)} + \frac{1-\alpha}{N} \qquad \text{…(2.4)}$$

where L(j) is the total number of neighbors of node j.

2.2.1.5 Degree centrality

Degree centrality(DC) is based on the concept of straight links in the networks. The degree centrality of a node is the summation of all the direct links associated with node *x*, as shown in Eq.(2.5).

$$DC(x) = \sum(\mathrm{dig}\ (x)) \qquad \text{…(2.5)}$$

where dig(x) is the degree of node x.

2.2.2 Metrics

In order to measure and evaluate the network characteristics, the following metrics are used in the social network analysis.

2.2.2.1 Circulation

These types of metrics are used to define the sharing features of the node in the network. These metrics include bridges, distance, density, Preferential attachment, and structural holes.

- **Bridge:** Bridge indicates the connection between the nodes that are part of the big network. There is zero overlapping neighbor(s) between the nodes.

- **Distance:** Distance indicates the path between the two nodes in the network. If one node is the direct neighbor of another node, the distance between the nodes is one. Other variations of distance are walk, path, geodesic distance, etc.
- **Density:** The network density is defined as the ratio between the total number of actual edges and the possible total number of edges in the network. A network is dense if most of the nodes in the network are connected through direct edges.
- **Structural holes:** Structural hole shows a critical role in information diffusion. The structural hole is defined as the vacant space between a node's neighbors, i.e., non-overlapping nodes exist among the neighbors. Structural holes allow the flow of information from one different network to another network that exists on either side.

2.2.2.2 Association

Association is defined as the tendency to establish a connection with the other nodes in the network. The user attempt to make a connection with the other nodes based on their interest and requirements. These include multiplexity, homophily, transitivity, proximity, and preferential attachment.

- **Multiplexity:** Multiplexity defines the total number of separate links between the two nodes in the network. In the real world, two nodes can be connected through various relations like friend and colleague, senior and neighbors, and many more.
- **Homophily:** Homophily is defined as the node's affinity to make the connection with the other nodes having similar attributes and behavior. Homophily is the main element in the social network to make a connection based on common interests.
- **Transitivity:** Transitive property is used to define the closure of a node in the network. For analyzing the 'friend of friend' connection, such a metric is used. If A likes B and B likes C, there is a chance that A also likes C.
- **Proximity:** Proximity refers to a person's propensity to establish a relationship with others based on closeness. If a person is in the workgroup of another person, there is a more probability of making a relation based on spatial proximity.

- **Preferential attachment:** Preferential attachment property support the richer get richer phenomenon. According to this, if users have a higher number of connections than other nodes, they have more chances to receive more relations than other users having bad relationships.

2.2.2.3 Segmentation

- **Clique:** Clique represents a closed group of users in which all the users are directly connected with each other without any intermediate node. It is a kind of complete graph in which no other can not be added without any importance.
- **Clustering coefficient:** Clustering coefficient refers to the degree through which nodes are inclined to make a cluster in the network. The global clustering coefficient defines as the division of closed triplet (three edges) to the total number of open(two edges) and closed triplet. The range of the clustering coefficient lies between zero and one. If there is no triangle in the graph, the clustering coefficient might be zero.
- **Cohesion:** Cohesion defines the degree of togetherness or inseparableness among the nodes in the network. A node is more cohesive if it is closely connected with pairs of nodes in the network. Similarly, a cluster is more cohesive if the most nodes are strongly tied with each other.

2.3 Review progression

A systematic process is carried out to perceive the various opinion leader identification techniques, as shown in Fig. 2.1. The entire process is categorized into four steps: the first step is the drafting of research queries, the second phase includes the exploration approach, the third phase consists of the norms to compile the whole study, and the fourth step is the plan for data collection.

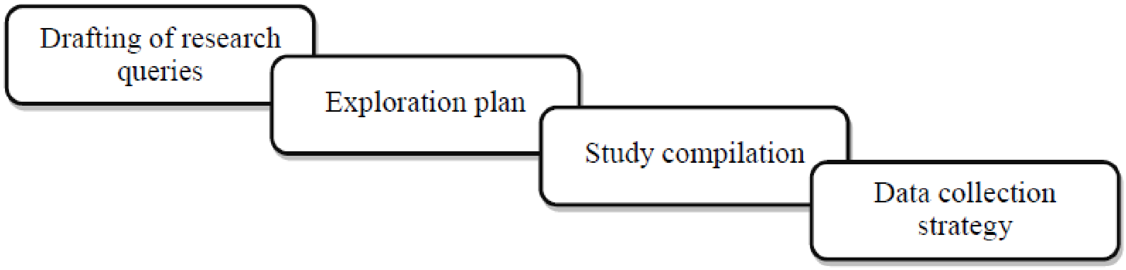


Fig. 2.1. Phases of the review process

2.3.1 Drafting of research queries

To systematically discover the substantiation opinion leaders and also identify their power, the following queries are recognized:

- **RQ1:** What should be the methods to find out the opinion leaders in the online social network based on computational intelligence techniques?
- **RQ2:** How can the accuracy of the methods improve?
- **RQ3:** What is the character and power of opinion leaders in information diffusion?
- **RQ4:** What are the techniques to aggregate the alliance of opinion leaders?
- **RQ5:** How to explore the applicability of opinion leaders in healthcare through the online social network?

2.3.2 Exploration plan

Exploration strategy includes organized and managed actions to determine the relevant and appropriate studies and innovations after identifying the RQs. So, few keywords such as opinion leader, online social network, computational intelligence techniques, power, healthcare, information diffusion are pulled out from the RQs. These terms are the root for the further research plan. Some closest synonyms like influencer, reputation, Twitter, etc., are also used for searching purposes. Besides, the combination or permutation of mentioned keywords is used for exploring the studies. Some boolean operations are also used for a more in-depth inspection.

Various research articles, thesis, dissertation, digital libraries, conference/Journal papers published by the reputed publishers/digital societies are studies in this researcher. Total 11 publishers/digital organizations like IEEE, ACM, Springer, Elsevier, Wiley, Taylor & Francis, SAGE, MDPI, Emerald, IET, and The MIT press are selected practically significant work. Various searching criteria like keyword, title, subject domain, and abstract are used for searching in these portals. Fig. 2.2 shows the total number of publications published under the various publications related to opinion leaders.

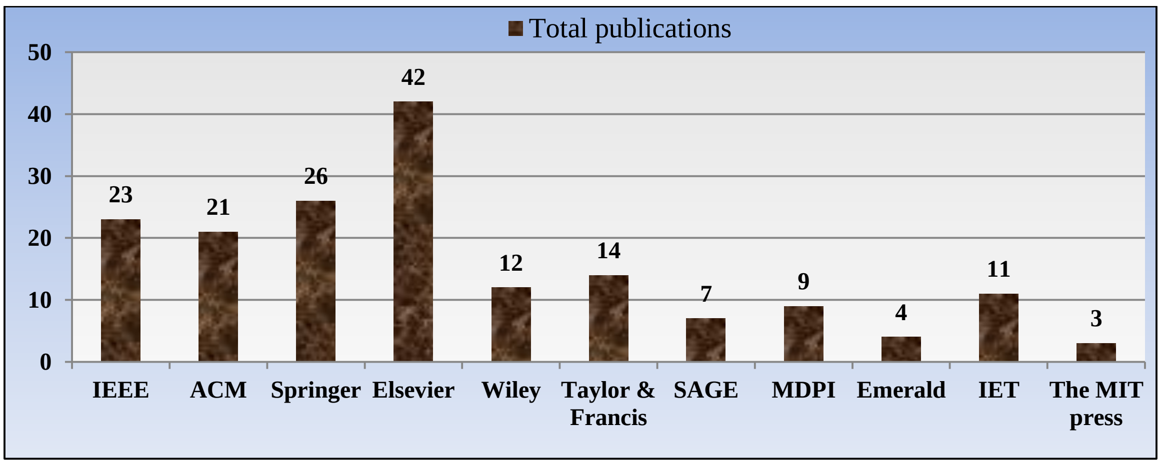


Fig. 2.2. Total number of publications under the various digital publications

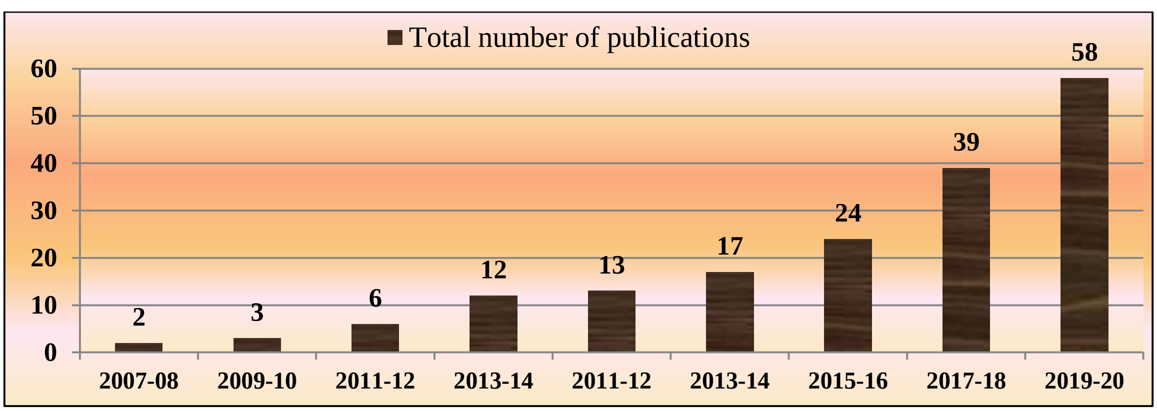


Fig. 2.3. Year-wise publication details

The selected studies are bounded into the time span, i.e., the period for searching the previous researches is a decade starting from 2009 to concluded till 2020. Fig. 2.3 demonstrates the year-wise total number of studies related to opinion leaders in the online social network.

2.3.3 Study compilation

It is challenging and critical to identifying those articles/findings that are incredibly appropriate for the work. In this phase, an addition-rejection policy is applied to find out the relevant publications.

2.3.3.1 Addition policy

Addition policy includes the following guidelines:

- All the research articles related to opinion leaders in online social network
- All the survey, techniques, and frameworks that are related to opinion leaders.
- Websites related to online social network-related datasets
- All the measures that are used for comparative analysis related to opinion leaders.
- All the articles that are depicting the applicability of opinion leaders in healthcare.
- All the articles demonstrate the power of opinion leaders for information diffusion.

2.3.3.2 Rejection policy

Rejection policy includes the following guidelines:

- All the studies related to opinion leaders but have no novel work.
- All the studies are related to non-online opinion leaders.
- All the studies are other than the English language.
- All the studies are not related to our RQs and do not fulfill our objectives.

2.3.4 Data collection strategy

This phase includes the identification of those articles that maps to the particular RO. The extraction of valuable information from the specific theme is very challenging and complex. It consists of mining techniques, approaches, theories, datasets, outcomes, limitations, publication year, author credentials, journal/conference's reputation, and future scope of the particular article.

In a nutshell, a total of 172 articles are identified by applying the mentioned keywords on the 11 different digital publishers. After identifying and removing the similarity in the paper pattern, only 89 papers are selected. Further, after considering the addition-rejection policy, only 35 articles are chosen. Next, after reviewing the ROs and checking the novelty, uniqueness, and results of the papers, only 22 articles are finally selected. Besides, seven cross-reference articles are also considered for the study. Thus, a total of 29 (22+7) papers are found extremely suitable for further analysis and investigation. So in the literature survey, all these papers are deeply investigated along with their positive and negative features. Fig. 2.4 illustrates the flow chart of the entire review process in detail.

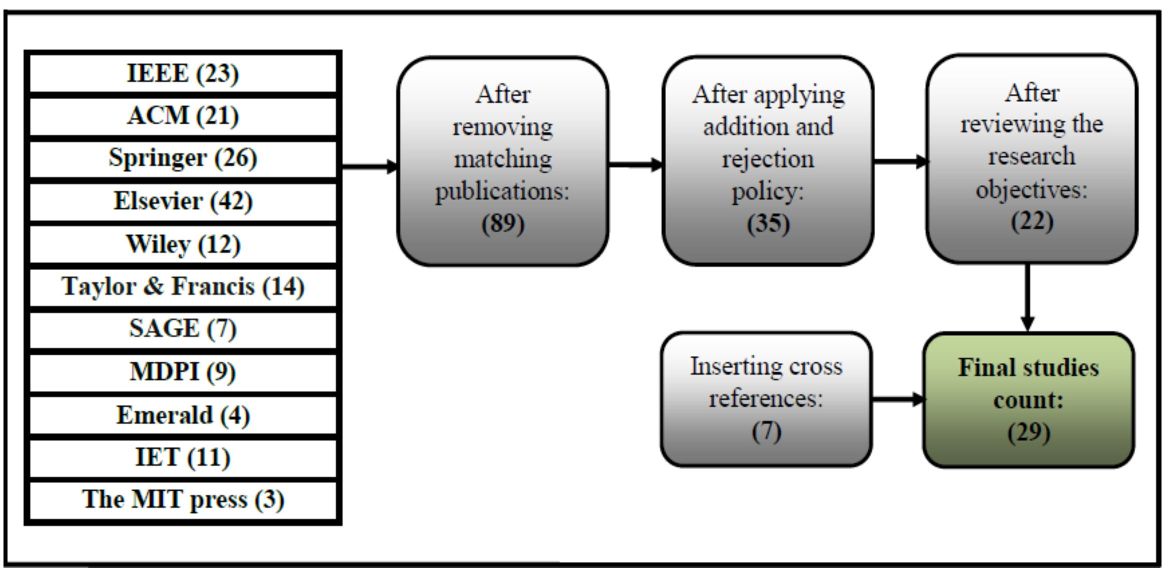


Fig. 2.4. Flow chart of the review process

2.4 Literature survey

In the present era, with the innovation of Web 2.0, there is the immense importance of social networks in human life. Lots of researches have been carried out to uncover the opinion leader in a social network. The researchers studied various approaches based on numerous parameters such as trust models and metrics, the total number of tweets, fuzzy logic, belief network, game-theoretic model, clustering, and text mining, and many more. In [88], researchers used the enhanced version of the Advogato trust metric that found the healthiest path from one user to another based on the maximum flow method. This method also found reliable users and avoids unreliable users from contacting the user's group. In [47], a framework is proposed that analyzed the user behavior by using their text content and responding to the responses. After investigating these attributes, the user ranked based on their longevity and centrality.

In [89], a novel approach is proposed based on user opinion, text mining, and relationships with other network users. In [90], a time-based propagation model is proposed in which the authors defined a decay function that dealt with the user's influence and constructed the chain of action specific impact of the leader. In [91], researchers identified the domain of sensitive opinion leaders in online review communities in which the domain of the contents helped to classify the opinion leader. In [92], the authors proposed an approach based on diffusion speed and the maximum growing number of adapters. After analyzing those factors, ranked the user based on their high sociality. In [93], researchers discovered the opinion leader in web-based stock message boards using clustering and sentiment analysis. In this model, users are clustered based on their posts that they posted on the stock board. Further, they applied sentiment analysis to finding the opinion leaders. In [94], investigators designed a new model called LUCI (Longitudinal User Centered Influence), which took the user's connection as input and categorized them into four different clusters. Next, classification techniques are applied to classify the users into followers and opinion leaders. In [53], researchers proposed an InfluenceRank algorithm that measured the importance of the content that a blogger posted in the blogosphere. The blogosphere's effectiveness is also compared to other blogs. After measuring the content, ranked the opinion leader in the blogosphere. In [95], investigators addressed a new algorithm, called OLFinder, that computed each user's status score based on their content and popularity in a certain area. Next, the opinion leader ranked based on the probability of being an opinion leader.

Next, in [96], the authors suggested a new approach in the dynamic social network using promising network methods and tried to resolve the influence of overlapping issues in the community structure. In [97], researchers deliberated a new-fangled framework that computed each user's total trust value and found those users having the highest total trust value in online communities. In [98], inventors proposed a dynamic opinion rank algorithm called POLD (Positive Opinion Leader Detection) algorithm in Chinese news remark. They structured the whole network in the three levels where level one consisted of news, level two indicated a single-level network, and level three showed a single topic user network. Next, a mapping among all three levels applied and identified the most promising users in the network. In [52], an algorithm is proposed that used the opinion leader to resolve the cold start problem in the recommended system. In [99], a k-means clustering-based approach in a big social network is suggested. They considered the instant of posting a message to decide the weight of an edge and generated a post-follow relationship to construct a social network. Next, they identified the network's overlapping community structure and identified the opinion leader in each community. In [100], an innovative approach is proposed that considered key opinion leaders' discovery using a subspace learning algorithm in which common subspace identified for various modalities. This algorithm used the modality-consistent harmonized discriminant embedding (MCHDE) to find out the low-dimensional discriminative from the social network. In [101], researchers proposed an approach based on signal spreading in the social network. A signal spreading is a process in which each node affects its neighbor via multi-relational vector association, and the importance of a node identified using the iterative matrix technique. In [102], investigators proposed an innovative method depending on closeness that indicted the relationship between the adjacent and non-adjacent nodes based on different interaction time and delays. The user with the highest closeness value is considered the opinion leader in the network. In [103], a novel hypothesis proposed that opinion leaders are topic-specific, i.e., opinion leaders have more authority and influence in a particular field. They used various topic models to determine those users who had more knowledge and control on those topics. Finally, the opinion leaders discovered based on maximum information spread.

Besides the mentioned literature, some researcher has different outlooks and ways of network formation to recognize opinion leaders. In [104], the researcher stated that the various human life opportunities made an opinion leader. They proposed that if a user received first-hand information from recognized sources, then no need to rely on any other social media. The gathered information itself sufficient to make them opinion leaders. In [105], the authors defined the role and explored opinion leaders' influence on the diffusion process of famous mobile games. They experimented that if an opinion leader promoted the game-related topic, the diffusion rate of those words is relatively too high as compared to advance by the user-generated content. In [106], the author proposed a unique K-means framework that includes three phases; the first phase identified the list of opinion leaders having a maximum local reputation, the second phase espoused an innovative active game model to discover the locally Pareto-optimal community structure and the last stage used the opinion dynamics model for creating the growth of the sentiment matrix. One of the applications of the opinion leader in Smart Grid that consists of 'Prosumer' who not only uses the energy but also generates the energy used by other grids or consumers. In [107], the researcher addressed the new approach that used the individual prosumer's energy density to find leader prosumers in Smart Grid. They also discovered a dynamic game model to identify the prosumer-community group structure and experimented on various Smart Grid datasets to validate the performance of the system. Table 2.1 demostrates the entire literature review in brief.

Table 2.1. Summarized review of literature papers

S. No	Ref.	Dataset	Mechanism	Benefit	Limitations
1	[108]	Online forum	Text mining and social network analysis based new method has suggested in which positive and negative opinions formed. Opinion leaders selected based on their total positive opinions.	Able to extract opinion automatically; useful for business modeling	Limited text database; Only applicable for a static data set
2	[89]	Reviews of Apple's iPhone	Proposed a model in which the connections among the user's cluster detected. Social behavior of the users analyzed based on these connections. Finally, the text mining technique used to identify opinion leaders.	Simple and easy to implement; Helpful in recommendation system and trend analysis.	Applicable only for online communities; Lack of validation database.
3	[88]	Epinions dataset	An extensive Advogato trust metric proposed that encourages the recognition of reliable user concerns with other users. Next, recursively diffuse the limit of designated users all through their system. Finally, it exhibits the capacity-first maximum flow technique for finding the most grounded way appropriate for finding a set of valid users.	Manage to get authorization in the network; Provide a solution for the recommended system.	Trust metric is not valuable for worldwide express connections; Access only the binary relations in the network.

4	[92]	Synthesized dataset	Explored the opinion leader based on the best-promoting decision, dissemination speed, and the most extreme total number of adopters utilizing age model. Besides, the opinion leaders with high sociality are considered for quick dissemination based on the reenactment result-users with high disjointing centrality considered as the most extreme aggregate number of adopters.	Useful for marketing and business purpose; Effective for indirect recommendation	Adopters information needed; Limited centralities used for calculations; Viable characteristics of the products neglected.
5	[47]	Educational blogs dataset, Parent-child forum dataset	A mixed approach proposed contained user connections, user behavior, user prominence, and time of responding recognized. The opinion leader discovered based on four features.	Very helpful in online forums and blogs; high execution as far as centrality.	Covered only a few online forums and blogs; Only topic-specific constraints considered.
6	[91]	Amazon.com dataset	A model is designed to detect domain-sensitive opinion leaders in online forums and networks using a model. For the ranking purpose, identify the customer's domain of interest, user's specialty, and transition probability matrix.	Helpful for product dissemination; Suitable for inclusion and authority.	Limited scope; The user's interest needed for evaluation purposes.

7	[90]	Facebook dataset, citation network dataset (Google scholar & DBLP)	Characterized an exponential time-decay formula to gauge leaders' impact and build the chains of leaders' action-specific influence. At that point, next, manufacture the global chains by normalizing every conceivable user's activities.	The proposed model is helpful to find an influence path; High accuracy and precision.	Too complicated to understand; not suitable for all datasets; Lack of parameters to chose the path.
8	[93]	Web-based Chinese stock Forum dataset	User's activities measured based on the text submitted on the board. Next, applied the clustering algorithms on the user's information and set of appropriate opinion leaders identified. Further, the critical opinion leader discovered based on the sentiment analysis to find the real price movement patterns.	Useful for marketing and business strategies; Easy to implement.	Only compared the outcomes with PageRank centralities; Utilized only for online blogs and forums.
9	[95]	Synthesized dataset	The proposed algorithm detects the central issue in the domain and then measure the leadership score of each user in the area. The opinion leader discovered based on the highest leadership score.	Better average precision and precision-recall; Easy to understand	Do not consider the total number of retweets; The output is topic dependant.
10	[101]	Douban.com	An algorithm proposed to identify the node status based on importance matrix signaling spreading, i.e., how a node penetrated the entire network. The opinion leader selected	Converged in lesser iterations; High accuracy and efficiency.	Only page rank used for compared results; Few features used to manage the

			based on the node status and importance.		importance matrix.
11	[97]	Epinions dataset	An innovative framework depends on trust in which total trust value (TTV) and opinion leader selected found on maximum TTV.	Better in terms of in-degree, out-degree; Hybrid IO-degree method,	Less accuracy; Uses limited trust metrics.
12	[96]	Mobile01 Forum	An efficient procedure, D_OLMiner, suggested for identification of opinion leaders in a robust social network. Next, a network rising technique proposed to build a dynamic social network to recognize the network community, and unravel the impact covering the issue and decrease the total calculation time.	Highly efficient; Implemented dynamic social network; Better impact flooding	Compared the outcomes with few metrics; Applicable only for online forums
13	[109]	Mobile01 Forum	Addressed the innovative OLMiner algorithm used the behavior and standard neighbor connection to reduce the total number of generated candidate and choose the list of opinion leaders.	Overcome overlapping influence problem; High efficiency and scalability.	Not suitable for all social network datasets; Limited similarity metrics used.

14	[110]	turnbackhoax.id	Implemented two approaches: Edge weighting and Centrality weighting; Edge weighting used to find the opinion leader by counting the total number of tweets and retweets. Centrality weighting is used to assigned weight to each centrality for finding the opinion leader more accurately.	Suitable for rumor spreading in the network; identify the importance of each edge; Highly accurate.	Limited centrality measures and relationship used; Applicable only on a few social networks.
15	[100]	KOL dataset using Instagram API, Synthesized dataset	Opinion leader identification is considering a multimodal skilled job. An innovative subspace learning algorithm addressed to discover the low-dimensional discriminator from social media data to select key opinion leaders.	Better subjective and quantitive outcome; Presented multimodality and high-dimensionality characteristics.	Highly multifaceted; straight-direct modification is not sufficient for the whole informational collections.
16	[102]	Synthesized dataset	Opinion leader detection dependent on the new closeness based calculation that incorporates a distinctive kind of collaboration among the users. The calculation additionally relies upon the communication time and postponement of nodes in the system.	Progressively productive and precise; designed updated independent cascade model; show more power than betweenness centrality.	Highly complex; Only implemented on a static dataset.

17	[103]	Sina micro-blog dataset	The opinion leader's selection is based on topic-sensitivity constrained. At first, the spread properties and the user's topic pertinence qualities estimated and haphazardly accept by 10% of the user. Opinion leaders classification dependent on the time of data diffused and the number of tainted nodes.	Profoundly effective; Improved Independent Cascade model; the lesser number of seed nodes needed with limited data spreading time.	Pre network structure and node information required; Due to the assorted variety of topics, not appropriate to work with a variety of datasets.
18	[111]	Chinese Sina BBS	A PageRank-based algorithm called Hybridrank proposed considering topic sensitivity and temporal features. Topis sensitive analysis used to find groups, while temporal feature analysis used to obtain the influence of opinion leaders over time.	High accuracy over PageRank algorithm; Easy to implement	Only considered link, topic and temporal characteristics; applicable for a small dataset
19	[106]	Facebook, Twitter, Google+	Proposed a framework based on a discrete-time dynamic model. Opinion leaders identified based on densely connected components having a similar opinion vector.	High Average cluster quality; More accurate and robust dynamic model.	Complex in nature; Not suitable for overlapping communities.
20	[112]	Twitter	Defined a unique component milestone centrality includes interest and exclusivity on some topics by participating users. The	Proposed innovative centrality measures; Easy	Experiments performed only on few topics;

			users with the highest milestone centrality elected as opinion leaders.	to understand; No need to review relationship.	Milestones needed to validate the results.
21	[113]	Local Motor communit y	Proposed a method included degree, median, and near centrality as multi-features attributes. The benefit indicator and cost indicator used to improve the accuracy of the method. Opinion leaders are selected based on the highest multi-features characteristics.	Robust; Straight-forward and easy to understand.	Lack of the degree of promotion; Only used for static datasets.
22	[114]	travel communit y dataset	Opinion leaders identified in virtual travel communities based on construal, content influence, and action to measure the influence on consumer's decisions.	Used SNA in virtual travel community; measure opinion leader influence efficiently	Only applicable for few datasets; Consider few parameters for identification.
23	[115]	Higgs Boson data from Twitter	Proposed an approach based on in-degree and out-degree to identify the opinion leader. Next, rank the user by their BC value. It also implemented the standard Louvain method for community detection.	Simple and easy to implement; Less complex.	Only a few centralities considered; Limited scope; Used static dataset.

Now, few studies are addressed related to rumor spreading that supported the foundation of this research. In [116], an innovative model is proposed based on the mean-field concept to find the decisive threshold value in the communities. The user worked as a mobile agent with some probability responsible for the inter-group long-range movement and disease-free clusters. The proposed approach also determined the transmission effect of the disease in the network. In [117], A memory-based model is suggested that explained the impact of memory over time. Also, they have used a few attributes along with memory rate that has more influence on rumor transmission. According to [118], a unique 8-phase ICSAR model is proposed that covered different features to find rumor dissemination. They have also considered the eight reasonable elements like information interest, rumor goal, users' inclination, trust scale on social media, transmission rate, augmented component, block score, and specialist influence to evaluate the rumor. In [119], researchers addressed a modified trust-based SIR model that defined a few mathematical equations representing the dynamics of the SIM model in the network. They also stated that the trust component could reduce the network population by limiting an evaluative threshold to control rumor transmission momentum. In [120], the analyst investigated that followers are the leading player responsible for rumor/antitumor diffusion or prevention. So they studied the nature of users who already have faith in rumors and can influence others by their strong beliefs. As per [121], researchers established a new Susceptible–Hesitating–Affected–Resistant(SHAR) model that evaluates user's behaviors against rumors along with local and universal rumor concern equilibrium. They also computed the reproduction and sensitivity of rumors towards parameters change for the model. They have also analyzed the three different human viewpoints for experimental purposes. In [122], the authors investigated the importance and briefness of weak ties in the rumor spreading. They argued that the role of weak ties is not much powerful for rumor propagation, but the diffusion rate depends on the identification of weak links. They have also defined a probability based on weak-tie dependent function.

In [123], analysts proposed two transmission channels; One is used for point-to-point transmission, and another channel is used for global rumor transmission. A modified SIR model also addressed using the mean-field formula. To authenticate the proposed approach and verify the impact of rumor spreading, a geometric function along with simulation has performed. In [124], a novel rumor transmission model has been proposed based on the time interval and non-linear procedure. Initially, the basic reproduction number calculated based on the next-generation matrix. Next, the researchers analyzed the rumor stability and factors needed for stability survival using experimental simulation. According to [125], the researchers developed an innovative SIR-based SKIR model that integrated rumor and anti-rumor-related facts. They have also identified the active factors that stimulate the people to accept the rumors and the anti-rumors using the game theory approach. They have also found the internal and external user characteristics that support rumor's impact using regression analysis. They have also considered the user's attitude and dynamic behavior for rumor propagation.

Thus, various approaches are put forward for rumor controlling and spreading from the user's and social perspective. Most of the methods are based on epidemic models, user's response time, distributing speed, rumor transmission rate, rumor's content, user's behavior, attitude, and opinions. The major pitfall with these approaches is the lack of user's trust, reputation, and content accountability. Although, few strategies also worked on the discussed issues but either fulfilled only a few limitations or could not demonstrate the power of opinion leaders in rumor prevention.

2.5 Research gap and limitations

After performing the profound, systematic literature on the selected studies, most concealed facts have been identified. Research gaps have been discovered as follows.

- **Less optimized outcome:** It has been observed that almost all the techniques produced the optimized effect according to specific datasets. Once they are applied to some other dataset, their performance might be reduced or not produced the optimized results. In this study, all the approaches work well with almost all types of datasets and give the same performance. The outcomes are also compared based on accuracy, precision, recall, F1-score error rate, and execution time with other standard SNA measures.
- **Discover only global opinion leaders**: Identifying opinion leaders is an extremely challenging task for the large degree of the dataset. All the previous investigations only identified the opinion leaders at the universal level, while in this research, the opinion

leader is not only recognized at the global level but in each community also at the local level.

- **The limited set of datasets:** In the previous studies, mostly analysis performed on the online forums, the blogospheres, and review datasets. Only a few of them used the social network datasets to identify the opinion leaders. This research incredibly supports all types of datasets because only domain-specific user attributes are needed to calculate the user's centralities.
- **No balancing between the local and global search**: Almost all the previously discussed approaches do not formulate the proper balance between the local and global tracking. Local search somewhere follows the greedy path and tries to find out the solution around its present solution. The global search focused on the best possible optimal solution and attempted to cover potential remoteness within the search space. In this work, two nature-inspired metaheuristic algorithms are proposed to balance the exploration phase and exploitation phase within the search space.
- **Highly complex structure:** Most of the procedures are very complex and included a complicated logical structure, mathematical functions, derivatives, and equations. Also, algorithms needed too much computational time for the execution. The proposed approaches' rational design is straightforward and understandable that does not include any complex and inconsistent formation.
- **No parameter control:** Almost all the approaches used the fixed parameters derived by using pre-tuned algorithm-reliant observations. The suggested strategies are performed well as the number of users increases gradually due to the dynamic nature of the control parameter.
- **Usage of data mining approaches:** Most of the studies used traditional data mining techniques such as clustering, text mining, classification, and many more to discover the opinion leaders. Although these techniques are well-received for most problems, yet data mining techniques are not much more appropriate for social networking applications due to the obscured data structures and the high volume of inappropriate data.

- **No community detection algorithm:** Generally, all the methods utilized the various domain-specific attributes to identify the universal top-N opinion leaders in their research. There are no methods which identified the communities in the network concerning opinion leader. Therefore, few techniques have been designed to find the communities in the network to discover the local opinion leaders.
- **No rational decision making:** Most of the approaches used conventional methods and functions based on graph theory and complex networks to identify the opinion leader. There is a lack of rational decision-making to make logical and sensible decisions. The main benefits of rational decision-making are to choose the best possible impartial solution over alternatives. Thus, a unique amalgamation of social network analysis and game theory is addressed to produce the best possible solution.
- **No concept of the synergic coalition:** In the synergetic partnership, various users participated in a group and conveyed compensation obtained from their marginal payoff and guaranteed a consistent and robust association formed. Till now, no approach considered this kind of group formation to find the opinion leader. The first time such type of composition has been proposed for opinion leader detection.
- **Limited graph representation:** Mainly, all the discussed approaches do not consider the latent features of the graph; they used the network's characteristics such as clustering coefficient, centralities, density, structural holes, and many more, as the necessary component for analysis purpose. On the other side, GNN models are highly potent, operative, and active on graph province data. Analyzing a graph is a very complicated task compared to different kinds of data such as image, audio, text, etc. GNN models can represent the structural graph information due to the extensive influence of Graph embedding and CNN.
- **Lack of feature vectors:** Most of the approaches used the node's attributes and social network properties to access opinion leaders. Adjacency or connection matrix is derived for depicting the predetermined graph in 2-D format. In contrast, GNN uses the network embedding representing the node and edge features into low dimension space while protecting the network knowledge and formation. The primary advantage of

feature vectors is determining the hidden dependencies among the network nodes and links.

- **No application for rumor prevention:** Till now, no study focuses on opinion leaders' applicability in healthcare by preventing COVID-19 rumors. The only few studied addressed that how opinion leaders are effective in promoting the right physicians. In this study, a new approach is proposed that identifies opinion leaders' power to stop COVID-19 related rumors that eventually impact mental and physical health.

  Thus, the critical observations swing the direction from classical approaches to modern approaches that significantly use the modern computational intelligence techniques and reach the new height of standard. This study put one step forward in the direction of novelty, innovation, and knowledge in the field of computer science.

2.6 Chapter summary

In this chapter, an organized and managed literature review is presented that profoundly identified the multiple significant studies for the opinion leader identification in the online social network. The whole review plan is divided into four phases. The first phase involved the development of research queries. Next, a systematic study plan has been described that selects only the relevant studies related to opinion leaders from the various publications. Further, an addition-rejection policy is also applied that removes the total number of publications based on prescribed guidelines. Finally, few plans are used to collect the data. In the entire study, 172 papers have been identified for learning, but eventually, only 29 articles are found appropriate for the research. Thus, this chapter presents relevant and up-to-date literature about opinion leaders significantly.

Chapter 3

Opinion leader identification using the nature-inspired meta-heuristic optimization algorithms

Various novel nature-inspired algorithms have solved many real-world complex problems due to their efficiency and agility. Such a multifaceted problem requires a very linear solution that can optimize the outcomes effectively. Traditional models are not much capable of solving such types of monotonous and complicated issues. Nature-inspired meta-heuristic algorithms are simulating the behavior of the organism living in the real world. The leading reason behind this popularity is collective behavior, followed by all the agents together. All the agents interact with each other, show the self-organizer behavior, and present some intelligence to achieve the optimal solution. This chapter covers the two nature-inspired meta-heuristic algorithms to identify opinion leaders in online social networks effectively. Section 3.1 explains the first approach based on a social network-based modified firefly algorithm to find local and global opinion leaders in the social networks. Section 3.2 covers the social network-based whale optimization algorithm in detail. The comparative analysis of both the algorithm is also discussed in the section. Section 3.3 wraps the summary of both the algorithms.

3.1 Firefly algorithm-based approach

A novel and advance approach is proposed to determine the local and global opinion leaders in the community. Initially, a customized Louvain method based on the network's modularity gain is suggested to discover the communities in the social network.

3.1.1 Community partitioning algorithm

A Community is a sub-organization or gathering of users in the organization that speak to specific organization highlights. In an interpersonal organization, it is an exceptionally pivotal errand to distinguish networks. For instance, communities can be a bunch of individuals who follow a similar religion, a gathering of creatures living in a similar geological district or organization of creator coauthor who deals with a comparable subject, and so forth. The structure of the modified Louvain community detection algorithm is presented in Algorithm 3.1.

Algorithm 3.1: Modified Louvain community Detection

Input: Directed social network graph G = (V, E)

Output: The hierarchical set of communities G'.

Steps:

1. G'={}, the empty set of communities
2. node $n_i \in$ G, put node n_i in its own local community C;
3. while (j ≠ max_no_of_node) do
    a. for all node n of G
    b. position node n in its neighboring community C' including its own which have the maximized modularity gain Mj.
    c. end for;
4. end while;
5. if (new_modularity>old_modularity)
6. G' = the network between communities of G;

    else
7. stop;
8. end if;

A modified Louvain method is addressed to find the communities in the online social network. The main element of the process is a greedy method-based hierarchical clustering. Modularity gain of the network points to the compactness of the edge in the network. The value of modularity gain exists in the interval -1 and +1. The addressed Louvain algorithm is divided into two steps. In the initial stage, every node is fit into the community based on density. In the later stage, every node's modularity gain is calculated, associated with the adjoining cluster having the maximized modularity gain. The modularity gain can be computed as shown in Eq.(3.1)

$$Ɱ = \frac{1}{2m}\sum_{xy}\left[A_{xy} - \frac{k_x k_y}{2m}\right]\delta(c_x c_y) - (1-C_x) \qquad \ldots(3.1)$$

In the above equation, communities of node x and node y are c_x and c_y, respectively. The worth of function δ is one if both x and y belong to the same community; otherwise, it is 0. The variable A_{xy} represent the weight between node x and node y. the factor $\frac{k_x k_y}{2m}$ represent the expected number of nodes between node x and node y where k_x and k_y are the degrees of node x and node y, respectively. 2m indicates the total weight of the network. The value of modularity gain Ɱ lies between 1 and -1. In the modified method, it has been observed that the clustering coefficient C_x of node x is also affected the modularity of the community. As soon as the clustering coefficient gradually increases, the probability of the node being the portion of the same community also increases and vice versa.

3.1.2 Firefly algorithm's variation in the social network

Xin-She Yang instigated the Firefly algorithm in late 2007 at Cambridge University [126]. The basic idea of the firefly algorithm is based upon the flashing comportment of fireflies. A variation of the mentioned algorithm is proposed for the social network to discover the opinion leaders. In the interpretation of the said algorithm, the user behaves as firefly, and the user's attractiveness is proportional to the prominence of the user. Therefore, the following guidelines would propose by our algorithm:

- Initially, all the user in the social network is treated homogeneous user regardless of their gender.
- The attractiveness (β) of the user (Firefly) is proportional to the prominence (brightness) of the user, and as the centrality (distance) decreases, prominence about a user also decreases due to triadic closure and homophily. The user having lesser

prominence, try to find out a user having more prominence about a particular domain. If the user could not find the one having high prominence, the user stops the user's search.

- The prominence of the user is determined by the user's centrality (C).
- Rank the user based on an attractiveness score.

3.1.2.1 Attractiveness

The attractiveness (β) of the user is proportional to the prominence observed by the adjacent users. The attractiveness (β) of the user i within the community is explained as shown in Eq.(3.2).

$$\beta = \beta_o e^{-\gamma C^2} + P_i \qquad \ldots(3.2)$$

Where β_o is the attractiveness at C=0 and P_i is the prominence value.

3.1.2.2 Progress

The progress of the user i towards the more attractive user j, can be computed using Eq.(3.3).

$$x_i^{t+1} = x_i^t + \beta_o e^{-\gamma c_{ij}^2}(x_j^t - x_i^t) + \alpha_t \epsilon_i^t \qquad \ldots(3.3)$$

Where the second fraction of the equation exists due to the attraction and the third fraction of the equation indicate the randomization process with α_t as a randomized parameter and ϵ_i^t a vector of random numbers. When the value of γ parameter is zero, it is equivalent to the Levy flight firefly algorithm.

3.1.2.3 Search optimality in firefly algorithm

The vital component of any of the nature-inspired algorithms is the searching optimality that indicates how efficiently the algorithm obtains the desired result. The firefly algorithm also exhibits the two searching strategies: exploitation and exploration to find the local and globally optimal. In the proposed approach, the exploration is gained by randomization, which includes the user's random search in the social network to find the other user having more attractiveness. Therefore this searching approach is beneficial to find the optimal opinion leader in the global extent. The appropriate amount of balance is needed between randomness and the global search for achieving the desired result. If the randomness is too high, it may be possible that the algorithm converges very soon or neglect some local minima of the subgroups; hence proper tuning is required.

For implementing the exploitation, information and knowledge about the local region, the total number of users, each user’s neighbor, the distance between the users, and many more attributes are required. Besides this information, some other knowledge related to subgroup shapes such as convexities, gradients, and past processes records are also needed to find the local minima. The pragmatic study from the observation state that exploitation tends to enhance the convergence rate of the algorithm; on the other hand, exploration tends to decrease the algorithm's convergence rate.

The harmonizing between exploitation and exploration is dependent upon the nature of the network and its surroundings. Landscape-based optimality includes the complete information regarding the whole network, total no of users, centrality, in-degree, out-degree, clustering coefficient, and many more so that an optimal solution can be found at local as well as global level with a lesser amount of exertion concerning time and number of iterations. Although there is no proper guideline for landscape-based optimality, this synchronization mainly depends on concrete landscape-based optimality. In the proposed approach, both exploitation and exploration are used to find the local and global optimum minima.

3.1.3 Parameters setting

For implementing the social network-based firefly algorithm, the heuristic control parameters are initially set up with the best setting for the proposed research. Initially, some random values are considered for the parameters and compared the result accordingly. In this research, all the parameters are not heuristic; some static parameters also exist, such as the size of the network, the centrality of the node, and the degree of trust. Due to the large size of the network, only 40% of the population is considered for the training set. The rest 60% of the population is used for validating the parameter's value in the research.

In this work, attractiveness (β_0), light absorption coefficient (γ) as prestige, randomize parameter (α), the total number of users in the network (n), and the maximum number of iterations (i) are heuristic parameters. The combination (n*i) is appropriate for obtaining the solution space of the problem. If the value of (n*i) factor is very large, it is a privileged probability to obtain the better value for all the parameters; but due to memory, time, and resource constraints, only 5000 nodes considered of the 'small Slashdot' dataset to hold the calculation of parameter within time. The value of all the parameters is analyzed using the linear model Analysis of Variance (ANOVA) approach [127]. Initially, three groups are selected in which; the first group consists of 1000 users that process over 100 iterations, the second group consists of 2000 users that process over 50 iterations, and the third group consists of 500 users that process over 2000 iterations. For filling the entire entries in the ANOVA table, the additive centrality is considered of each user for calculation.

The ANOVA model consists of Degree of Freedom (DF), Mean Square (MS), Sum of Squares (SS), F-statistic, and P-test values as shown in Table 3.1. Only those parameters are considered as a statistically optimal parameter for which the value of $P \leq 0.05$ with 95% confidence intensity.

Table 3.1. Values of heuristic parameters using ANOVA

<table>
<tr><th rowspan="2">n*i</th><th colspan="3">Parameters</th><th rowspan="2">Degree of Freedom (DF)</th><th rowspan="2">Mean Square (MS)</th><th rowspan="2">Sum of Squares (SS)</th><th rowspan="2">F-statistic</th><th rowspan="2">P-test</th></tr>
<tr><th>A</th><th>β₀</th><th>Γ</th></tr>
<tr><td rowspan="3">(500*200)</td><td>0.1</td><td>0.1</td><td>0.01</td><td>2</td><td>45893</td><td>2674981</td><td>25.87</td><td>0.000</td></tr>
<tr><td>0.25</td><td>0.25</td><td>0.1</td><td>2</td><td>46348</td><td>2745896</td><td>32.51</td><td>0.000</td></tr>
<tr><td>0.5</td><td>0.5</td><td>0.15</td><td>2</td><td>52872</td><td>3876403</td><td>59.07</td><td>0.003</td></tr>
</table>

	0.75	0.75	0.2	2	35789	1784975	39.52	0.000
	1	1	0.25	2	27841	879423	17.93	0.000
(1000*100)	0.1	0.1	0.01	2	45893	2758106	39.05	0.000
	0.25	0.25	0.1	2	56489	3917393	45.92	0.000
	0.5	0.5	0.15	2	62958	4728491	65.20	0.005
	0.75	0.75	0.2	2	60013	4187928	52.69	0.001
	1	1	0.25	2	41433	2711942	21.72	0.000
(2000*50)	0.1	0.1	0.01	2	16784	1078349	24.97	0.000
	0.25	0.25	0.1	2	27538	2187592	36.80	0.001
	0.5	0.5	0.15	2	38222	4007828	48.53	0.004
	0.75	0.75	0.2	2	32989	3335713	27.41	0.000
	1	1	0.25	2	29483	2967485	15.36	0.000

From the above statistical result, it is concluded that the optimal value for the heuristic parameters attractiveness (β_0) is 0.5, the value for light absorption coefficient (γ) is 0.15, and the value for randomizing parameter (α) is 0.5.

3.1.4 Complexity of firefly algorithm

The complexity of most of the nature-inspired algorithms is uncomplicated and simple. The firefly algorithm consists of one outer loop that iterates over the total number of maximum iteration m and two inner loops that iterate over the total size of network n. so the worst-case complexity of the algorithm is (n^2m) [128]. Although, the complexity of the algorithm is linear if the iteration is very far above the ground and n is relatively near to the ground. The computation of the algorithm also engrosses the computation of attractiveness value that is a linear function ω, involves the computation of the degree of trust and centrality of the entire node in the network and the complexity of this fraction is ($n\omega$) that is linear. Therefore, the proposed algorithm's overall complexity is Θ ($n^2m + n\omega$) $\approx$ O (n^2). The pseudocode of the algorithm is as follow:

Algorithm 3.2: modified Firefly Algorithm

Input:

1. Generate an initial population of users u_i(i=1,2, 3…n)
2. Initial light absorption coefficient (γ) as prestige.

Output: list of users having the highest attractiveness

Steps:

1. Compute the prestige of each user based on the degree of trust.
2. Set all the heuristic parameters attractiveness (β_0), light absorption coefficient (γ), randomizing parameter (α) derived using the ANOVA model.
3. Evaluate initial attractiveness of the user u_i.
4. Measure user's progress toward other user u_i.
5. while(iteration<=max_iteration)
   {
6. for (i=0 ; i<user_population_size; i++)
7. for(j=i ; j<user_population_size; j++)
   {
8. if (Pi >Pj)
9. Update the user's attractiveness via $e^{-\gamma}$
   }
10. Update current user attractiveness
11. iteration++;
    }
12. return the list of users with their attractiveness.
13. end;

The firefly algorithm is initiated with a set of users called population and defines the light absorption coefficient (γ) as prestige for each user in the network. In this case, the user's behavior is not similar to the flashing behavior of the firefly that releases the light to attract the other fireflies; therefore, the initial value of the light absorption coefficient (γ) is computed using the degree of trust that one user in the network achieves. The degree of trust can be calculated with the following Eq.(3.4).

$$T_{xy} = f\,(d_{xy},\, d_{yx}\,,\, d^{z}_{xy}\,,\, d^{z}_{yx}\,,\, r_x,\, r_y) \qquad \text{...(3.4)}$$

Where d_{xy} is the degree of trust expand by user x on user y, d_{yx} is the degree of trust expand by user y on user x, d^{z}_{xy} is the degree of trust that is supposed to be suggested by user x to user y on user z, d^{z}_{yx} is the degree of trust that is supposed to be indicated by user y to user x on user z. r_x and r_y are the reputations of user x and user y, respectively. The trust can be a direct trust (DT), indirect trust (IDT), and recommended trust (RT) as shown in Fig. 3.1.

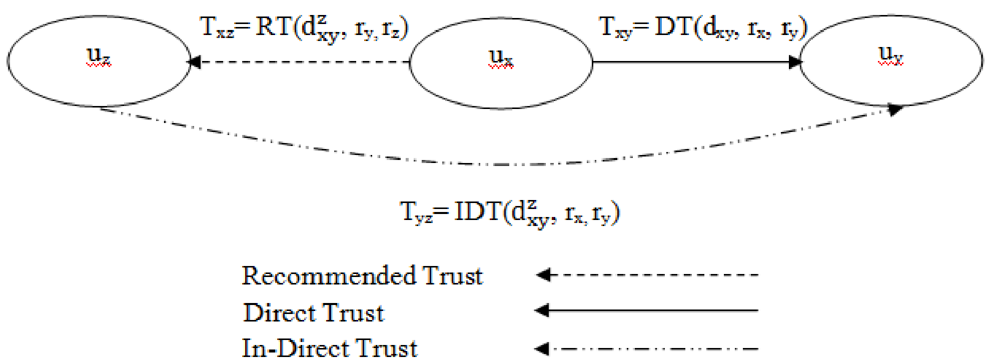


Fig. 3.1. Trust representation between the users

Further, each user attempts to search for a user with more attractiveness that can be computed with the help of the user's prominence. There are two iterations in the mentioned algorithm; the first iteration for the outer users and the second iteration for the current user whose prominence value is compared with the other users' prominence value. Suppose a user searches for a user with higher attractiveness. In that case, the first user updates its knowledge about the user's attractiveness and socializes this knowledge to its all dimension. Similarly, all the users pursue the identical procedure and associate the experience to its entire neighbor in all dimensions. At last, identify the top-N users who have higher attractiveness in the social network. The flow chart of the proposed algorithm is shown in Fig. 3.2.

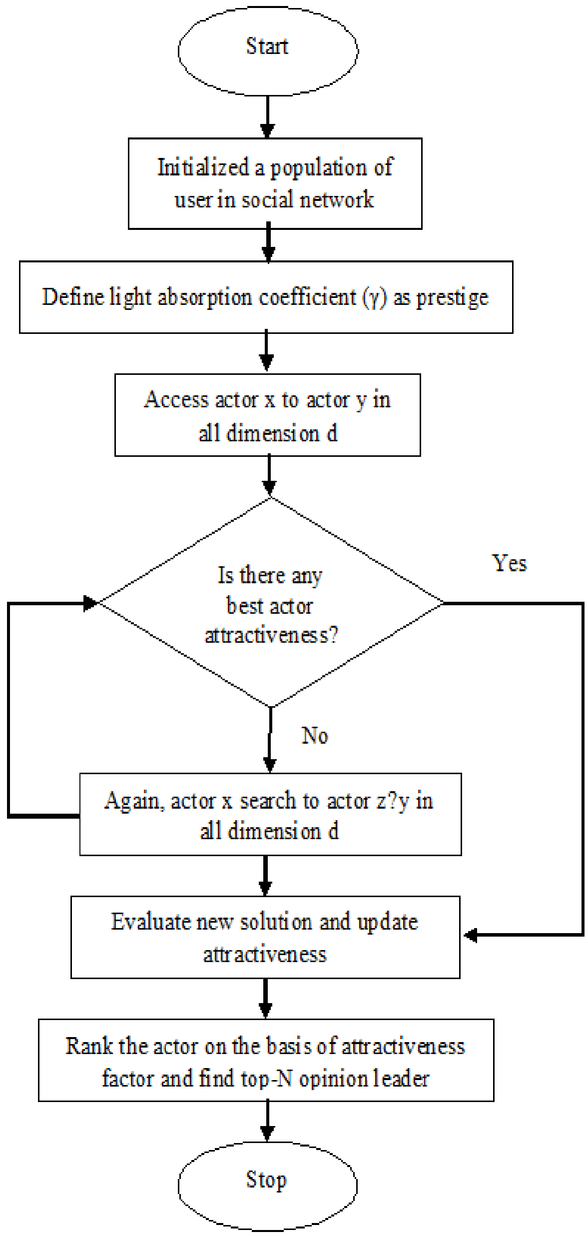


Fig. 3.2. Firefly algorithm flow chart

3.1.5 Why does the firefly algorithm preferred?

The main feature of the firefly algorithm that makes it so efficient to identify the opinion leader in the social network as follows:

- The firefly algorithm is a nature-inspired swarm intelligence-based heuristic algorithm in which multiple agents interacted with each other and solved the global optimization problems. The central concept of the firefly algorithm is based upon the brightness and attractiveness of the firefly. As soon as the distance between the users changes gradually, the user's attractiveness factor is also updated in the same proportion, and the whole population of the network is automatically divided into multiple subgroups. In each subgroup, all users move around the local optimum. Once the local optimum of all the subgroups has been measured, the most excellent global optimum solution can be established.
- This algorithm's second main virtue is that as the whole population is separated into numerous subgroups, fireflies permit to find of local optimal simultaneously in each community. Therefore, as the network's population amplifies, there is no effect on the computation time to find the local optimum.
- The third advantage of the firefly algorithm is that the control parameter, light absorption coefficient (γ), can be controlled as the iterations in execution to swift and speed up the chances of converges. The main benefit of this strategy is that once the result converges, the iterations can be discontinued and locate the optimal value for the control parameters.

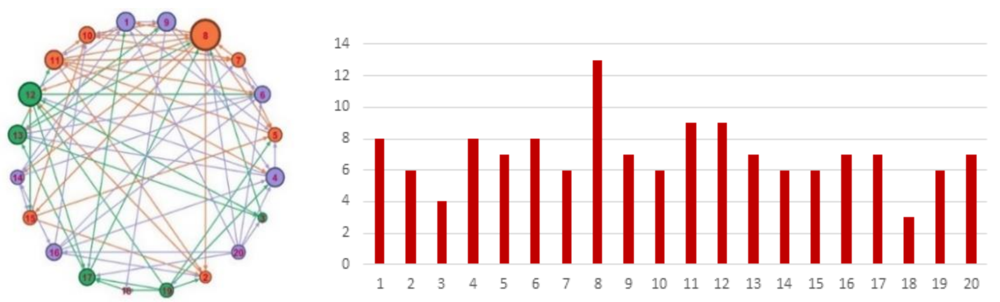
3.1.6 Experiment and evaluation

In this segment, the total number of local and global opinion leaders evaluated and identified by the proposed algorithm that can be implemented on synthesized and real data sets. In the beginning, first, the datasets are described and then compared the result with the standard measurement that is to be used to find opinion leaders in the online social network.

3.1.6.1 Datasets

3.1.6.1.1 Synthesized dataset

A synthesized dataset has used for the experiment. The dataset has a total of 20 nodes and 70 edges, as shown in Fig. 3.3(a). The network density is 7.00, and the degree of each node is shown in Fig. 3.3(b).



(a) (b)

Fig. 3.3. Structure of (a) Synthesized Social network, and (b) Node Degree distribution

According to the proposed approach, a modified Louvain community partitioning algorithm is applied and found the entire four communities for the dataset. In the next step, the proposed modified firefly algorithm is involved in the social network to compute each community's attractiveness to find out the local opinion leader according to their attractiveness, as shown in Table 3.2. The same algorithm is also applied for the whole network structure to identify the opinion leader globally. Once each user's attractiveness is computed, arrange the user according to their attractiveness and find the top-N (=5) number of global opinion leader in the network as shown in Table 3.3. The results are also compared with other SNA measures.

3.1.6.1.2 Real dataset

The proposed algorithm is also implemented on a real dataset that included the 'small slashdot' datasets, a network of friends and foes, as shown in Fig. 3.4. The dataset has

13182 nodes as users, and a total of 34621 edges represent the relationship (friend) between the users. There are 76.7% users are friend relationship and remaining is foes relationship. The density of the network is 5.1981. for analysis purposes, the same procedure is applied for this dataset as used for the synthesized dataset.

Fig. 3.4. Structure of slashdot social network

In the real dataset, there are some missing relationship values between the users exist. So, the arithmetic mean is used of the entire node's degrees for substantial those missing values. Initially, the total degree of trust is measured of each user for measuring the prestige. Further, the modified Louvain community partitioning algorithm is applied to this dataset, and a total of 28 community structures are identified. The identification of the community also depends upon the type of network. The user's attractiveness also relies on the landscape of the network that includes the overall knowledge about the network. In this dataset, the in-degree, out-degree, and total degree distribution of the entire nodes are shown in Fig. 3.5.

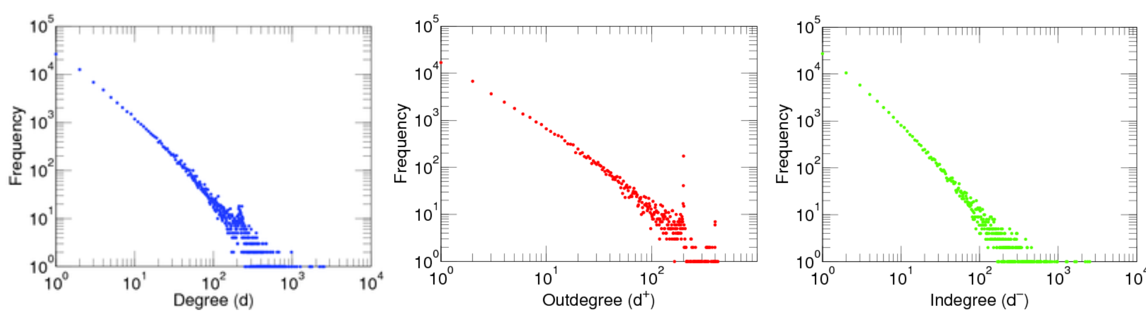


Fig. 3.5. Degree, Out-degree, and In-degree distribution of network

3.1.6.2 Results

The value of all the heuristic parameters is computed for the dataset and applies the firefly algorithm. The firefly attractiveness for the top opinion leader in every community is shown in Table 3.4. The top-10 global opinion leaders identified by the proposed approach and other SNA measures are also shown in Table 3.5.

Table 3.2. Top-3 Local opinion leader based on various SNA measures and proposed firefly algorithm in each community for the synthesized dataset

Local Comm-unity	Node id	DC	Node id	BC	node id	CC	Node id	PR	Node id	Firefly attractiveness
1	12	0.473684	12	0.063464	12	0.655172	12	0.062541	**12**	**0.2182500**
	2	0.315789	19	0.035791	2	0.575758	19	0.045186	**2**	**0.1195585**
	15	0.315789	2	0.022086	15	0.575758	15	0.043607	**15**	**0.1188108**
2	8	0.684211	8	0.142888	8	0.76	8	0.086868	**8**	**0.3923120**
	1	0.421053	17	0.045134	1	0.633333	1	0.055526	**1**	**0.1807533**
	13	0.368421	1	0.034884	20	0.612903	17	0.050955	**17**	**0.1531570**
3	11	0.473684	4	0.048933	4	0.633333	11	0.061513	**11**	**0.2081808**
	4	0.421053	11	0.042203	11	0.633333	4	0.056041	**4**	**0.1837085**
	5	0.368421	16	0.033431	5	0.612903	16	0.050074	**5**	**0.1510719**
4	6	0.421053	9	0.057866	6	0.633333	6	0.055904	**6**	**0.1818417**
	9	0.368421	6	0.041133	9	0.59375	9	0.051139	**9**	**0.1522720**
	14	0.315789	14	0.020969	14	0.575758	14	0.043487	**14**	**0.1193333**

Table 3.3. Top-5 Global opinion leader based on various SNA measures and proposed firefly algorithm for the synthesized dataset

Node id	DC	Node id	BC	Node id	CC	Node id	PR	Node id	Firefly attractiveness
8	0.684211	8	0.142888	8	0.76	8	0.086868	**8**	**0.392312044**
12	0.473684	12	0.063464	12	0.655172	12	0.062541	**12**	**0.218250008**
11	0.473684	9	0.057866	4	0.633333	11	0.061513	**11**	**0.208180898**
4	0.421053	4	0.048933	11	0.633333	4	0.056041	**4**	**0.183708542**
6	0.421053	17	0.045134	6	0.633333	6	0.055904	**6**	**0.181841780**

Table 3.4. Top Local opinion leader based on various SNA measures and proposed firefly algorithm in each community for real dataset

Comm-unity	Node id	DC	Node id	BC	Node id	CC	Node id	PR	Node id	Firefly attractiv eness
1	8	0.025036	8	0.050489	8	0.399739	8	0.003896	**8**	**0.600012**
2	757	0.013656	757	0.013203	237	0.362853	757	0.002864	**757**	**0.512457**
3	822	0.020256	822	0.021471	822	0.353719	822	0.00405	**822**	**0.491772**
4	898	0.021925	898	0.022510	898	0.364136	898	0.004114	**898**	**0.324512**
5	190	0.030802	190	0.036551	190	0.374471	190	0.005791	**190**	**0.598897**
6	520	0.011683	520	0.013942	454	0.338434	520	0.002719	**520**	**0.356568**
7	394	0.017449	394	0.020450	394	0.337542	394	0.004004	**394**	**0.462458**
8	163	0.016918	163	0.021447	163	0.352867	163	0.003862	**163**	**0.495554**
9	522	0.029436	522	0.037023	522	0.375666	522	0.005496	**522**	**0.596478**
10	825	0.02974	825	0.031148	825	0.35607	825	0.005823	**825**	**0.594271**
11	834	0.014946	834	0.016314	935	0.356407	834	0.003184	**834**	**0.426680**

12	523	0.018436	523	0.026269	523	0.371432	523	0.003503	**523**	**0.345609**
13	617	0.016539	905	0.019068	617	0.35139	905	0.003451	**617**	**0.487541**
14	642	0.030726	642	0.033205	642	0.359018	642	0.006029	**642**	**0.595370**
15	936	0.031409	936	0.024885	936	0.396266	791	0.004299	**936**	**0.594602**
16	62	0.03103	62	0.027942	62	0.380075	62	0.00478	**62**	**0.602541**
17	162	0.018056	162	0.018683	106	0.371149	162	0.003867	**106**	**0.587458**
18	344	0.01358	344	0.015740	344	0.361629	344	0.002594	**344**	**0.452112**
19	644	0.018208	644	0.019741	644	0.346112	644	0.00382	**644**	**0.475556**
20	57	0.017677	57	0.021058	57	0.368772	669	0.003462	**57**	**0.482331**
21	184	0.031105	184	0.085150	184	0.385669	184	0.006851	**184**	**0.601245**
22	794	0.015932	794	0.014683	794	0.354271	794	0.002878	**794**	**0.548854**
23	706	0.000607	706	0.001062	706	0.262063	706	0.00025	**706**	**0.574845**
24	855	0.011532	855	0.014769	453	0.338365	855	0.003007	**453**	**0.554002**
25	142	0.020712	142	0.022020	142	0.361837	142	0.003955	**142**	**0.495822**
26	173	0.025871	173	0.031230	797	0.353928	173	0.00569	**173**	**0.521406**
27	813	0.015477	813	0.013895	813	0.346131	813	0.002878	**813**	**0.560002**
28	248	0.011228	248	0.012686	248	0.344107	568	0.002602	**248**	**0.485690**

Table 3.5. Top-10 Global opinion leader based on various SNA measures and proposed firefly algorithm for real dataset

Node id	DC	Node id	BC	Node id	CC	Node id	PR	Node id	Firefly attractiveness
936	0.031409	184	0.085150	8	0.399739	184	0.006851	**62**	**0.602541**
184	0.031105	8	0.050489	936	0.396266	642	0.006029	**184**	**0.601245**

62	0.031030	522	0.037023	82	0.390757	825	0.005823	**8**	**0.600012**
190	0.030802	190	0.036551	43	0.388648	190	0.005791	**190**	**0.598897**
642	0.030726	642	0.033205	625	0.387688	173	0.005690	**522**	**0.596478**
913	0.030726	173	0.031230	913	0.385872	522	0.005496	**642**	**0.595370**
82	0.030574	825	0.031148	184	0.385669	62	0.004780	**913**	**0.594829**
791	0.030498	62	0.027942	791	0.383302	791	0.004299	**936**	**0.594602**
825	0.029740	523	0.026269	74	0.381494	898	0.004114	**825**	**0.594271**
522	0.029436	936	0.024885	62	0.380075	822	0.004050	**791**	**0.594158**

In the mentioned tables, proposed algorithm results have been compared with the social network analysis standard measures. Each table includes six columns, and every column has two sub-columns; the first sub-column contains the node id and, the second sub-column contains the value for the particular measure. Table 3.2 and Table 3.3 are for the synthesized dataset to discover opinion leaders in local and global leaders, respectively, while Table 3.4 and Table 3.5 are for real datasets to find opinion leaders in local and global leaders, respectively. The last column in each of the tables displays the node's firefly attractiveness according to the proposed method. Further, in this investigation, it has been identified that there is a correlation between the light absorption coefficient (γ) and the network size. It is also analyzed that changes in the values of light absorption coefficient (γ) and the network size also affect the correlation coefficient (r). As the value of the light absorption coefficient (γ) updated, the correlation coefficient's value also changes, as shown in Fig. 3.6.

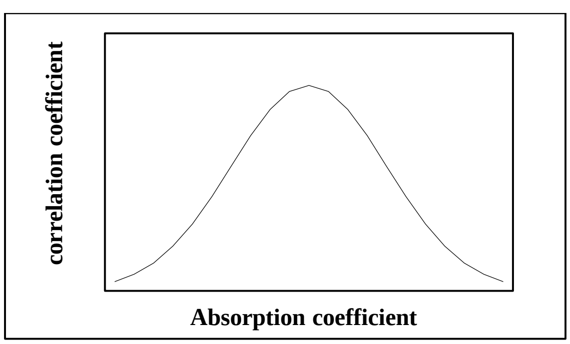


Fig. 3.6. Absorption and correlation coefficient relation

It is also observed that as the absorption coefficient increase initially, the correlation coefficient also increases, but later it decreases gradually as soon as the absorption coefficient increases. Hence, based on this fact, this can also analyze that the overall correlation coefficient (r) also changes as the total number of opinion leaders increases, as shown in Table 3.6.

Table 3.6. Correlation coefficient (r) and absorption coefficient (γ) relationship

absorption coefficient (γ)	r(N=10)	r(N=100)	r(N=500)	r(N=1000)	r(N=5000)	r(N=10000)
0.2	0.3250	0.3255	0.4021	0.4350	0.4822	0.5214
0.5	0.3511	0.3758	0.4832	0.5298	0.6235	0.7566
0.7	0.3355	0.3473	0.4521	0.4752	0.5214	0.6323

Furthermore, the experimental result also indicates that the total number of leaders identified in each community is also varied as the results proposed by our algorithm compared to the other methods used for the same purpose. It is observed that the total number of opinion leaders identified by each process in the above Tables demonstrated the percentage of the top 5% users discovered by each method out of the total users in the social network for the real dataset, as shown in Fig. 3.7.

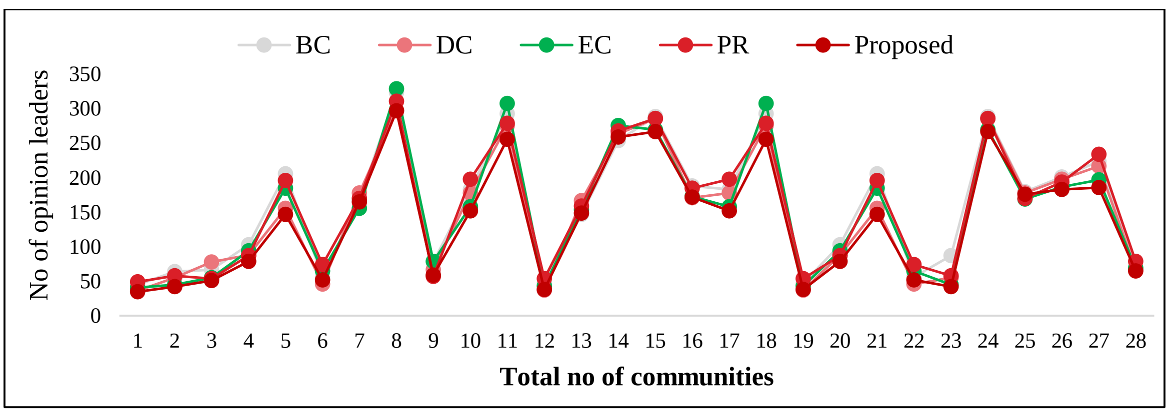
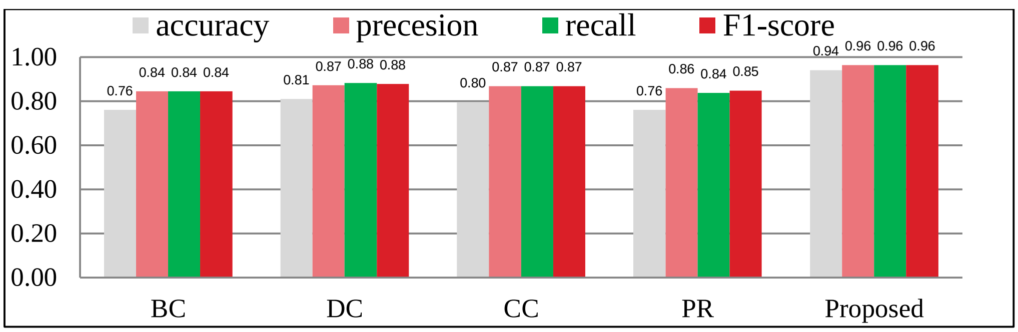
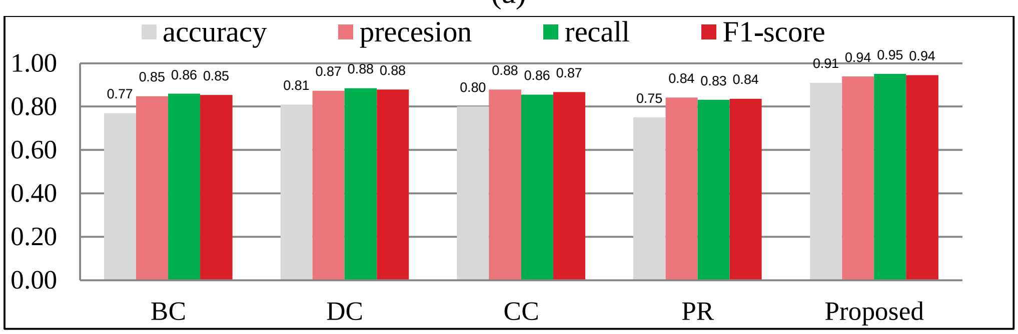


Fig. 3.7. Comparison of BC, DC, EC, PR, and proposed firefly approach for the top 5% users in each community for real data set

For the validation of the proposed method, the proposed research results are compared for accuracy, precision, recall, and F_1-score. It can be inferred that the proposed firefly approach yields better results as compared to other standard measures, as shown in Fig. 3.8. It can also be monitored that the total number of opinion leaders is found lesser in the synthesized and real dataset. Additionally, the total amount of dignified opinion leaders also varies as the heuristic control parameters' values revolutionize in a different province. The research's crux is to optimize the attractiveness of the user that persuades other assessments and perceptions about a particular object. The firefly algorithm is unsurpassed appropriate for the social network as the firefly's behavior matched with the behavior of the social network user. Moreover, It is also observed that the total computation time taken by our algorithm is also significantly less as compared to other SNA measures, as shown in Fig. 3.9.



(a)



(b)

Fig. 3.8. Comparison of firefly approach concerning the accuracy, precision, recall, and F_1-score for (a) real data set (b) synthesized data set

The modified Louvain method is helpful to uncover the consequential communities. The average running time and modularity value are compared between the original and altered Louvain method as shown in Table 3.7 and found that results are improved and optimized. Therefore, confidently it can be concluded that the proposed technique produces better outcomes as compared to other standard SNA measures.

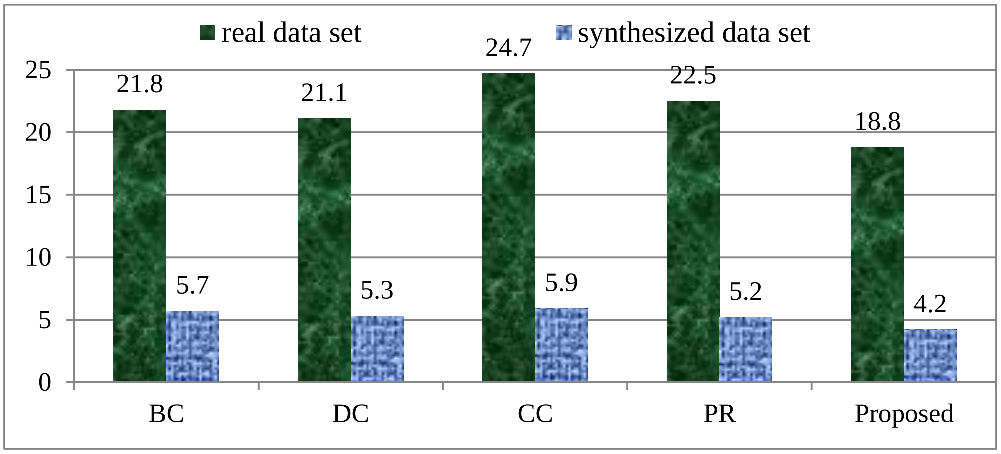


Fig. 3.9. Comparison of computation time for real and synthesized data set using firefly approach

Table 3.7. Comparison between proposed community detection approach and original Louvain method for real and synthesized data set

Attributes	Original Louvain		Proposed Louvain	
	Real	**Synthesized**	**Real**	**Synthesized**
No of nodes	13182	20	13182	20
Modularity	0.8522	0.542	0.7237	0.483
No of runs	100	10	100	10
Average running time (in sec.)	749	18	638	11

3.2 Whale optimization algorithm based approach

Another novel approach for community detection and social network-based whale optimization algorithm has been proposed. However, various techniques have been proposed to classify the communities based on multiple attributes [129]–[135]. First, the total number of communities is identified based on common neighbor similarities and the clustering coefficient. The whale optimization algorithm is discussed in the next section to determine the opinion leader in each community at the local level and in the social network.

3.2.1 Community detection algorithm

For identifying the node's community, the concept of common similarity between the neighbors of the node is used, i.e., neighbors adopted the common feature constituted the community. Some researchers [136], [137] found that the phenomenon of common similarity between the neighbors is similar to the clustering, yet there is a significant difference between these two concepts. Clustering is an unsupervised technique in which objects of the same type make a cluster based on some measure. So, the objects of the same cluster are similar while differing from the objects of another cluster [138]. The basic idea behind the common similarity of neighbors is to find the common things that like by the neighbor of a node. If a node has more common friends, i.e., friends of friends, it is apparent that they have more common choices and likes and have more chances to form a community. Therefore, neighbors' common similarity is one of the ways to find the clusters in the network [139]. The proposed algorithm assumes that each community must include at least five nodes to form a community. Although this is not a mandatory criterion for building a community that must have five nodes in the network yet, this condition has been imposed to reduce the total number of communities. In general, a community must have a k number of nodes to form a community, where k is a nonzero positive integer. Suppose the size of the dataset is too large. In that case, the network's modularity is also high, and the clustering coefficient of the network is a low positive value, then there is a probability that the network has more community of smaller size. A rigorous experiment has been performed for the different values of k on the same dataset and analyzed that the optimal value for k is five. Therefore, the condition has imposed that the communities must have at least five nodes in the dataset.

In this procedure, the edges are continuously removed from the network to find only those neighbors strongly connected with the node. According to the proposed algorithm, a community is formed based on the proposition that 30% of the community's users ought to be common neighbors among the user. The total number of neighbors $N(n_i)$ and clustering coefficient C of each node i are calculated in the network and calculated the total number of relations n_i among the common neighbor to identify the community. Now, the variables $N = n_i \cup n_j$, $T \leftarrow neighbor(n_i) \cap neighbor(n_j)$, and $W = neighbor\ (v_i) + neighbor\ (v_j)$ are computed where v_i and v_j are the end nodes of the edge e. Besides, the variable Z is calculated using Eq.(3.5).

$$Z = \left\{ \begin{array}{ll} \frac{N*U}{W*T} * C & ; \ \text{if}\ T > 0 \ \ and \ \ \frac{T}{U} < 0.3 \\ N * U & ; \ \text{if}\ T = 0 \end{array} \right\} \qquad \ldots (3.5)$$

Further, the edges are managed in ascending or descending order based on the Z value in Table M. Generally, Table M works as a container that stores the information about the edges in the network in chronological order. Now, the first edge e is considered and attempting to remove this edge from Table M. Once the edge is removed from Table M, the degree of other nodes is calculated as well. If the total number of the neighbor of the node v_i and node v_j more than zero, the edge is removed from the table. The same procedure followed continuously until the table empty. It has been mentioned earlier that each community must have at least five nodes. For ensuring this constraint, a minimum spanning tree for each community is made to ensure that it should have at least four edges. Another method for providing the same condition is to combine the entire subgraphs, having less than five nodes, to the last community and make one large community. The Pseudocode of the whole procedure is reviewed in Algorithm 3.3.

Algorithm 3.3: Community Detection Algorithm based on Neighbor similarity and Clustering coefficient (CDANsCc)

Input: network graph G = {V, E}

Output: total number of communities in a set C'= $\{C_1, C_2, C_3, ..., C_n\}$

Steps:

1. begin
2. for ∀ v ϵ V in G

   do

   $N(n_i) \leftarrow \{ e(v_m, v_n) \in E) \mid v_m, v_n \in neighbor(n_i) \}$ and compute the Clustering coefficient (C(v))
3. end for;
4. for ∀ $e(v_i, v_j) \in E$ in G

   do

   $T \leftarrow neighbor(n_i) \cap neighbor(n_j)$

   $U \leftarrow neighbor(n_i)\ U\ neighbor(n_j)$

   $N \leftarrow N(n_i) + N(n_j)$

   $W \leftarrow neighbor(n_i) + neighbor(n_j)$
5. if T >0

   if $\frac{T}{U} < 0.3$ then

   $Z \leftarrow \frac{N*U}{O*T} * C(v)$ and add $e(v_i, v_j)$ in M

   else

   $Z \leftarrow 0$

   end if;

   else

   $Z \leftarrow N * U$

   end if;

   end for

6. M ← sort the edges in descending order according to the Z value
7. do
8. if (degree(neighbor(v_i) ‖ neighbor(v_j)) > 0)
   remove the edge e(v_i,v_j) from D and insert into community C_i based on Z
   value
   end if;
9. while (M ≠ Ø)
10. for ∀ community C_i in C'
    do
    design the Minimum Spanning Tree (MST)
    if (cardinality(C_i) ≤ k)
    find the last community C_j in the network having a cardinality(C_j) ≤ k and
    $C_i \cap C_j = Ø$
    merge C_i with C_j and eliminate C_i from C'
    end if;
    end for;
11. end;

Further, the reputation r of the user measured using the optimization function. Therefore, initially, the objective function for each user is computed in the network.

3.2.2 Objective function

In a social network, the objective function O of each user can generate an active connection with another user. The objective function also represents the preference distribution of the user in the network. Whenever the position of the user changes, the value of the objective function updates accordingly. In the proposed approach, the objective function is defined based on BC, CC, DC, clustering coefficient C_i for a node, and distance $\vec{D}$ which is the difference between the user's current position and the best optimal position. The objective function is represented using Eq.(3.6).

$$O_i = \frac{\sum_{i=1}^{i=n} \sqrt{\frac{(BC_i.CC_i)^2}{DC_i}}}{\vec{D}} * C_i \qquad \ldots(3.6)$$

The main aim of the optimization function is to formulate the problem into a mathematical model. There are a set of known and unknown variables that manage the value of the optimization function. Therefore, an objective function is a mathematical function that one wants to maximize or minimize while preserving certain network constraints. In this approach, some limitations include the node's degree, clustering coefficient, each user's position must be known, and the network structure must be properly defined. Thus, there is a need to maximize the objective function by minimizing the distance among the users.

3.2.3 Node reputation

The reputation ***r*** of a user in the network represented a kind of credibility and prominence in the network [140], [141]. The reputation is also the benchmark for user's trust, authenticity, and security in the network. According to [142], reputation is what is said or believe about a person's past behavior or experienced by others. On the other hand, reputation is also defined as a combined measure of credibility or reliability based on ratings or recommendations [143]. In the proposed approach, a novel approach is adopted to measure the user's reputation. In this approach, the reputation of the user is directly proportional to the objective function. If the distance among the users is comparatively low, the value of the objective function becomes high. Eventually, the user's reputation might be lofty if the value of the objective function is increased. So, the reputation of a user is directly proportional to the objective function. If the network is analyzed deeply, it is found that the node having a higher reputation, containing more chances to become an opinion leader. There is a direct correlation between the user's reputation and various centralities measures. Different types of properties exist, such as triadic closure, homophily, strong and weak ties, triad, diad, and many more elementary properties [144] that also affect the user's degree of reputation in the social network.

3.2.4 Social network-based whale optimization algorithm

The phenomenon of whale optimization is proposed by S. Mirjalili, who explored the superiority of whale optimization using multimodal functions on various structural design problems [145]. The social network architecture is like a network in which each node represents the users, and each link represents the relation between the users. The whale optimization algorithm impersonates the humpback whale's behavior and uses the bubble net chasing techniques to prey as the enemy [146], [147].

3.2.4.1 Encircling prey

According to the whale optimization algorithm, the current whale location is considered the best solution to identify the optimal prey, and every whale changes its location towards the optimal location. The other whales also search for the optimal solution of the targeted prey simultaneously. For implementing the same model in the social network, each whale is considered a user. Another user can be regarded as prey for other's users but with a positive schism and approach. In 2-D space, the position of each user in the network can be represented using Table 3.8.

Table 3.8. Matrix representation of the user's position in the network

$$\begin{bmatrix} WU_{1,1} & WU_{1,2} & WU_{1,3} & \cdots & WU_{1,m} \\ WU_{2,1} & WU_{2,2} & WU_{2,3} & \cdots & WU_{2,m} \\ WU_{3,1} & WU_{3,2} & WU_{3,3} & \cdots & WU_{3,m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ WU_{n,1} & WU_{n,2} & WU_{n,2} & \cdots & WU_{n,m} \end{bmatrix}$$

Where $WU_{i,j}$ indicates the position of user i at location j in 2D space.

Initially, each user's current location is considered the best optimal location or near to the optimum. When the best optimal opinion leader is defined, a user attempts to update location towards searching the best optimal opinion leader with a higher reputation. During this process, the user activities and behavior can represent by Eq.(3.7).and Eq.(3.8).

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} . \vec{D} \quad \ldots(3.7)$$

$$\vec{D} = |\vec{B} \Theta \vec{X}_p(t) - \vec{X}(t)| \quad \ldots(3.8)$$

In the above equations, $\vec{A}$ and $\vec{B}$ are the coefficient vector, t indicates the current iteration, $\vec{X}_p(t)$ is the position vector for the best solution at iteration t, $\vec{X}(t)$ is the position vector, $\vec{X}(t+1)$ is the next position vector at iteration t+1, $\vec{D}$ indicates the distance between users at position $\vec{X}(t)$ and $\vec{X}_p(t)$, and Θ is an element by element multiplicative operator. It is also worth mentioning that in every iteration, the value of the vector $\vec{A}$, vector $\vec{B}$, and control parameter a, is calculated using Eq.(3.9), Eq.(3.10), and Eq.(3.11), respectively.

$$\vec{A} = 2\, a\, \vec{r} - a \quad \ldots(3.9)$$

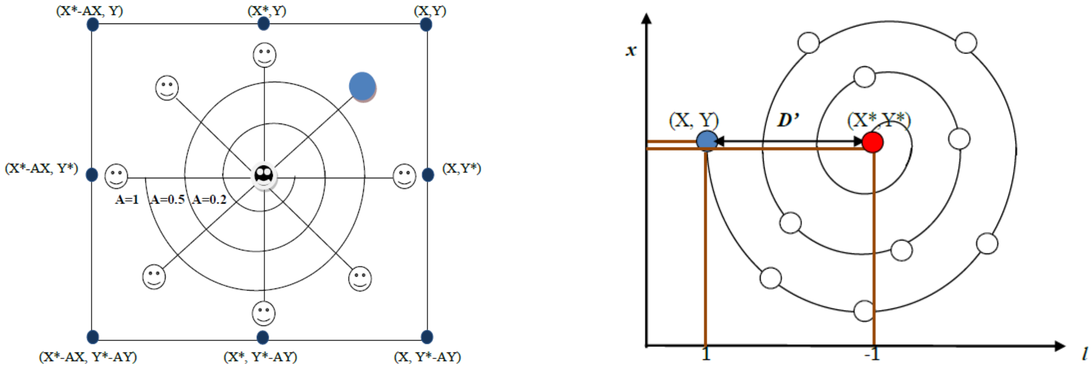
$$\vec{B} = 2\, \vec{r} \quad \ldots(3.10)$$

$$a = 2 - 2\,\frac{t}{t_{max}} \quad \ldots(3.11)$$

Where $\vec{r}$ is a random vector, a is a control parameter that varies between 0 and 2 over the iterations, and t_{max} is the highest number of iterations.

3.2.4.2 Exploitation phase: Bubble net attacking

Now, the bubble net attacking behavior for each user is discussed in the network. The bubble net attacking behavior of each user can be represented using shrinking encircling and spiral updating methods, as shown in Fig. 3.10.



(a) (b)

Fig. 3.10. Bubble net attacking method (a) Shrinking encircling method (b) Spiral updating position

3.2.4.2.1 Shrinking encircling position

In the shrinking encircling phase, each user's location can be defined anywhere between the initial position and the best possible optimum position prescribed by the control variable a, and its value varies between 2 and 0. Through intense observation, it is found that the value of the vector $\vec{A}$ also depended on control parameter a, and value of a vector $\vec{A}$ lies in the interval [-a, a].

3.2.4.2.2 Spiral updating position

In the spiral updating position activity, the distance between the user and the prey is initially calculated, as discussed earlier, for computing the optimization function. The user's position is at the location (X, Y), and the user's expected position is at (X*, Y*). Each user updates its position in the spiral's helix-shaped structure using Eq. (3.12) and Eq. (3.13).

$$\vec{X}(t+1) = \vec{D'} e^{bl} \cos(2\pi l) + \vec{X}_p(t) \qquad \dots (3.12)$$

$$\vec{D'} = |\, \vec{X}_p(t) - \vec{X}(t) \,| \qquad \dots (3.13)$$

Where $\overrightarrow{D'}$ indicates the distance between the user's current position and the prey user that demonstrates the best solution for another user, b is a constant value, and l is a random number having a range [-1, 1].

A user can move in a spiral updating position and shrinking encircling position concurrently. Therefore, whenever a user searches for another user having a higher reputation concerning other users, there is a probability p that the user can choose either the spiral updating position or the shrinking encircling position during the optimization. In the social network-based whale optimization algorithm, when the user updates their position, 50% either decide on the spiral updating position or the shrinking encircling position; because both the positions have an equal chance, i.e., 0.5, of being chosen. Therefore, 0.5 is only the probability of choosing any one position. This approach can be represented using Eq.(3.14a) and Eq.(3.14b).

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \, . \, \vec{D} \qquad p < 0.5 \qquad \ldots (3.14a)$$

$$\vec{X}(t+1) = \vec{D}' \, e^{bl} \cos(2\pi l) + \vec{X}_p(t) \qquad p \geq 0.5 \qquad \ldots (3.14b)$$

If the user does not follow either the spiral updating position or shrinking encircling position, they can move in random order, and movement can be represented using the Eq.(3.15) and Eq.(3.16).

$$\vec{D} = B. \, \vec{X}_{rand}(t) - \vec{X}_p(t) \qquad \ldots (3.15)$$

$$\vec{X}(t+1) = \vec{X}_{rand}(t) - \vec{A} \, . \, \vec{D} \qquad \ldots (3.16)$$

3.2.4.3 Exploration phase: Investigating for victim user

In this phase, the main essential factor of the exploration phase is randomization with the aim of random search by each user based on a few predefined network's parameters such as the total size of the network, overall in-degree and out-degree of each user, and homophily, as shown in Fig. 3.11. In this phase, a user can also move either in spiral updating, shrinking encircling, or in a random position. There should be proper steadiness needed between the random search and standard search. The algorithm might be converged if the randomness is too high in the network. Therefore, this phase is appropriate to discover the prey at worldwide altitude and appropriate balanced needed between global search and randomness.

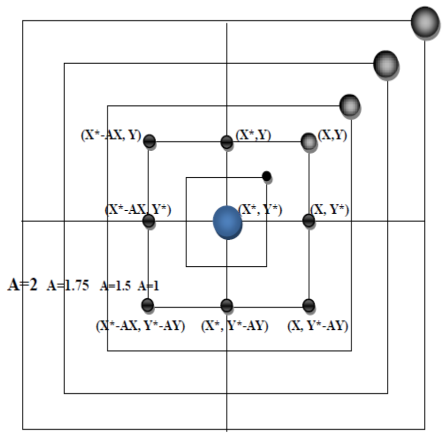


Fig. 3.11. Exploration phase of whale optimization

A flowchart and an algorithm for Social network-based whale optimization have been designed to explain the entire approach. A flowchart is the pictographic representation of an algorithm, and the user can quickly understand it. On the other hand, the algorithm is the step by step sequential solution of the problem. The overall structure of whale optimization can be represented using Fig. 3.12.

3.2.5 Complexity of the proposed algorithms

In this section, two algorithms are addressed; the first algorithm is designed to determine the opinion leader in the network at the community level and the global level, while the other is used to identify the communities in the social network. The complexity of both algorithms is not so diverse and is easily understood by others. In the community detection algorithm, initially, all the variables T, U, N, W, and Z are computed to set the network attributes. So, the complexity is linear; therefore, it needed only O(n) iterations to traverse all the network nodes. Next, the edges are sorted based on Z values which are computed in linear time O(m) and $m \leq n$. Further, each community's cardinality is measured to check whether the minimum spanning tree of a community has at least five nodes. For creating the minimum spanning tree, Kruskal's minimum spanning tree algorithm used that have complexity O(ElogV). As a result, the general complexity of the algorithm is O(ElogV) + O(m) + O(n) ≈ O(ElogV).

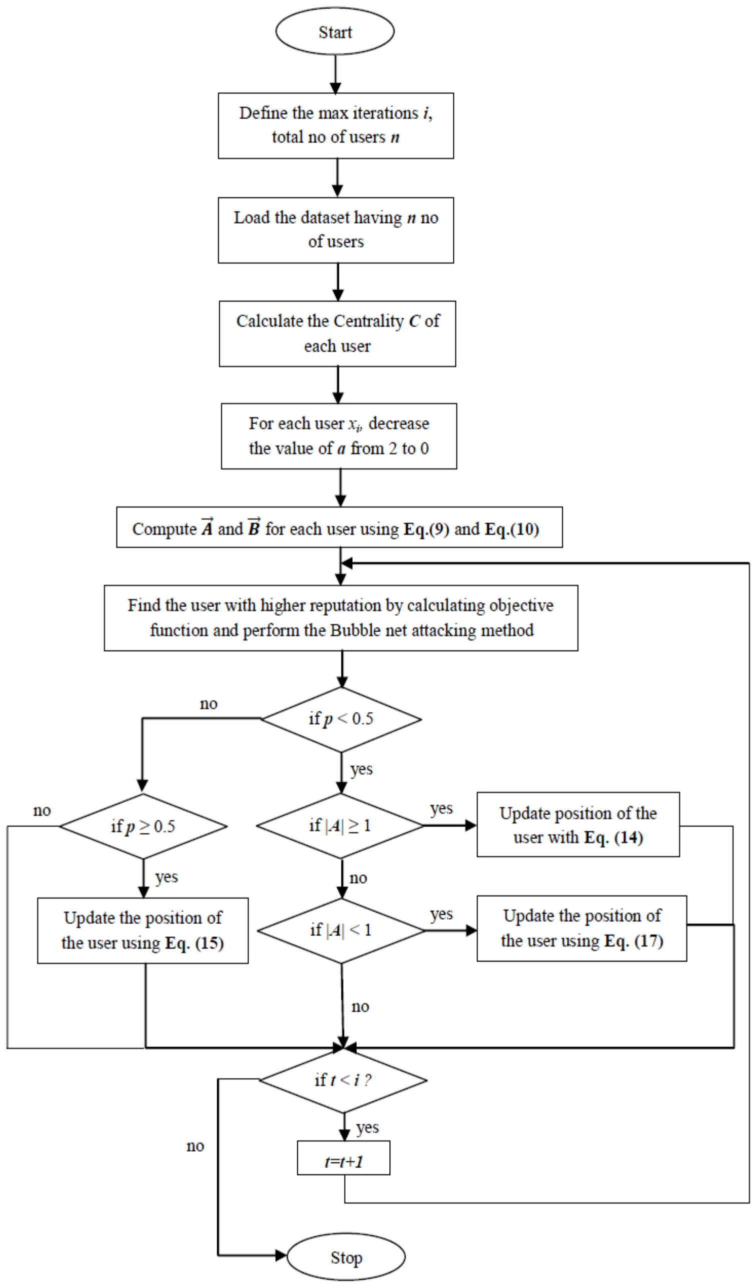


Fig. 3.12. Social Network based Whale Optimization Algorithm (SNWOA) flow chart

The pseudocode of the Social Network-based Whale Optimization Algorithm (SNWOA) is summarized in Algorithm 3.4.

Algorithm 3.4: Social Network-based Whale Optimization Algorithm (SNWOA)

Input: Social network Dataset having n users, total number of iteration i

Output: Optimized Distance between users

Steps:

1. begin
2. for j=1 to n

   do

   O[j] ← compute the initial objective function for each user using Eq.(3.6).

   end loop;
3. Evaluate $\vec{A}$ and $\vec{B}$ for each user using Eq.(3.9) and Eq.(3.10).
4. Re-evaluate the objective function of the best search user.
5. while (t < i)

   do

   if (p < 0.5)

   if (|A| ≥ 1)

   Update the position of the user using Eq.(3.14a).

   else

   Update the position of the user using Eq.(3.16).

   else

   Update the position of the user using Eq.(3.14b).

   end if;

   end loop;
6. t = t+1;
7. do step (3)
8. end;

In a social network-based whale optimization algorithm, the three centralities and distances are measured to compute each user's objective function in the network. This process's total computational time is O(4*n) ≈ O(n). Next, the coefficient vector $\vec{A}$ and $\vec{B}$ are measured that needed total O(n) computational time. Further, each user's position is updated either in a spiral updating position or in a shrinking encircling position, based on the reputation of other users, which required max (n) iterations. Hence, the entire complexity of the proposed algorithm is O(n) + O(n) + max(n) ≈ O(n), that is linear.

3.2.6 Why choose whale optimization algorithm in the social network?

A nature-inspired meta-heuristic whale optimization algorithm superlatively fitted for our proposed model due to the following reason.

- Various researchers recommend the whale algorithm for cracking the numerous problems [148]–[154]. This recommendation's motivating point is the social and intelligent behavior of whales that is somehow similar to human behavior. The sensing, learning, and communicating behavior of the whales are sturdy and robust compared to a human. Another most appealing feature of the whale is its bubble net attacking mechanism for probing the victim. In social network also, it is worth mentioning here that human is also eager to find a person having more reputation and prominence in the real world. Therefore, they try to follow the same pattern as supported by the whale.
- Another worth mentioning reason is that in the whale optimization algorithm that there is a control parameter a, having value between -2 to 2. So, this control parameter is responsible for the convergences of the algorithm, i.e., as the total number of users in the network increases gradually, the control parameters' values also change accordingly, and the algorithm terminates as soon as it converges.
- In the whale optimization algorithm, the optimal position of a whale for searching prey depends on another whale's relative position. Therefore, the position of the whale updated according to the uninformed, another whale agent. This movement's benefit is

that all the whales update their position optimally as the prey changes its position. In the social network, all users can also upgrade their position to identify the most significant opinion leader.

- Whale optimization algorithm works on exploration and the exploitation phase that enables the users to search the opinion leaders locally in the community and globally in the social network. As the number of users increases gradually, the whole network is partitioned into the communities, and the whale algorithm permits to find the opinion leaders in each community concurrently. Hence, the overall computational complexity of the algorithm does not transform.

3.2.7 Experiment and results

In this section, the suggested algorithm has been implemented on both the real and synthesized dataset. Initially, both the dataset details are discussed, and next analyzed the datasets' statistics such as modularity and clustering coefficient. A software, Gephi 0.9.0, as a tool, is used for analyzing the social network dataset. Python 3.6 is used for implementing the proposed algorithm. Finally, the results are compared with the other standard SNA measures [155] to show the effectiveness of the proposed algorithm.

3.2.7.1 Datasets

3.2.7.1.1 Wiki-vote dataset

The Wiki-vote dataset [156] is exploited as a real dataset with 7,115 nodes and 1,03,689 links between the nodes as represented in Fig. 3.13(a). This dataset contains the information about the votes that one user has given to another user to promote leadership on Wikipedia. The average clustering coefficient of the network is 0.1409, and modularity is 0.424.

3.2.6.1.2 Synthesized dataset

A synthesized dataset is also used to implement the proposed algorithms. The synthesized dataset is an undirected graph that includes a total of 100 nodes and 467 edges. The modularity of the network is 0.211, and the average clustering coefficient of the network is 0.1609. The network has a total of 17 strong triangles, and a direct link between two nodes represents the strong relationship.

(a) (b)

Fig. 3.13. Structure of dataset (a) wiki vote (b) synthesized

The modularity of a network is the main feature to determine the strength of the partition of a network into different modules and clusters. The modularity is often used to identify the communities in the network. A network's modularity is the variation between the fractions of edges that fall within that cluster to the expected fraction and the number of edges with the same node degrees circulated randomly. The range of the modularity lies between -1 to 1. The modularity of the dataset is 0.424, i.e., the communities in the network have, somewhere, dense structure and having more number of edges as compared to the expected number of edges if the network generated randomly with the same number of node degrees.

Clustering coefficient refers to the degree through which nodes are inclined to make a cluster in the network. The global clustering coefficient defines as the division of closed triplet (three edges) to the total number of open(two edges) and closed triplet. The range of the clustering coefficient lies between zero and one. If there is no triangle in the graph, the clustering coefficient might be zero. The clustering coefficient of the dataset is 0.1409, i.e., there are no more closed triplets present in the network. In the dataset, the modularity is high, and low clustering coefficient, i.e., the nodes having a high density in the communities but having the bare association with the other nodes in the different communities. It depends upon the network topology and dynamics that what type of nature is exhibit by the network. So, the modularity and clustering coefficient both represent the clustering behavior of the network in different aspects.

The proposed methodology is implemented on two datasets; the Wiki-vote dataset is a real-world dataset, while the other one is the synthesized dataset. The primary reason behind choosing the dataset is that both datasets follow the power law and scale-free properties of social networks [157]. A rich-get-richer model uses the probabilistic approach to implement the power law in a newly added node, with probability p, chooses a node randomly. The nodes expand new relations in proportion to how many they already have, i.e., some nodes end up with many more relationships than others. In the wiki-vote dataset, even the entire user’s vote and admin vote has equal weight; yet, the current admin has a higher probability to leverage the power of their network so that they can take control over the entire network. So, the wiki-vote dataset depicts the generalized structure of the social network. Some nodes have a higher degree in the synthesized dataset due to homophily and stable triangles, while other nodes have a lower degree. If a new node is added to the network, there is a higher probability that the higher degree node may grab the new node. Therefore, both the datasets exemplify the general structure of the social network.

The real dataset has some missing values, and for handling those missing values, the attribute mean value method is used in which the unknown values are filled by the attribute mean of the identified values. Next, the closeness centrality, betweenness centrality, degree centrality, and clustering coefficient for each user are calculated. The distance D between the users in the network is also measured. Further, each user's objective function in the network is evaluated and applied the proposed algorithm to find out the optimal minimum distance between the users. Whenever the distance between the users is changed, the objective function is evaluated based on centrality and distance value, as discussed in the previous section. The range of the objective function varies between [0, 1]. Different standard benchmarks functions [158] are used for optimization associated with a whale optimization algorithm to find the minimum optimized distance. The description of the optimization function is represented in Table 3.9.

Table 3.9. Standard benchmark optimization functions

Function name	Description	Range
Dixon-price function	$f(x) = (x_1 - 1)^2 + \sum_{i=2}^{d} i\,(2x_i^2 - x_{i-1})^2$	$x_i \in [-10, 10]$ for ∀ i =1, ..., d
Beale function	$f(x) = (1.5 - x_1 + x_1x_2)^2 + (2.25 - x_1 + x_1x_2^2)^2 + (2.625 - x_1 + x_1x_2^3)^2$	$x_i \in [-4.5, 4.5]$ for ∀ i =1, 2
Drop wave function	$f(x) = -\frac{1 + cos(12\sqrt{(x_1^2 + x_2^2)})}{0.5(x_1^2 + x_2^2) + 2}$	$x_i \in [-10, 10]$ for ∀ i =1,…, d
Three hump camel function	$f(x) = 2x_1^2 + 2x_1^4 + \frac{x_1^6}{6} + x_1x_2 + x_2^2$	$x_i \in [-10, 10]$ for ∀ i =1, 2
Bukin function	$f(x) = 100\sqrt{\lvert x_2 - 0.01x_1^2 \rvert} + 0.01\lvert x_1 + 10 \rvert$	$x_i \in [-10, 10]$ for ∀ i =1, 2
Matyas function	$f(x) = 0.26\,(x_1^2 + x_2^2) - 0.48\,x_1x_2$	$x_i \in [-10, 10]$ for ∀ i =1, 2
Easom function	$f(x) = -\cos(x_1)\cos(x_2)e^{(-(x_1-\pi)^2-(x_1-\pi)^2)}$	$x_i \in [-10, 10]$ for ∀ i =1, 2
Ackley function	$f(x) = -a\exp\left(-b\sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}\right) - \exp\left(\sqrt{\frac{1}{d}\sum_{i=1}^{d}\cos(cx_i)}\right) + a + \exp(1)$	$x_i \in [-10, 10]$ for ∀ i =1, …, d and a=20, b=0.2 and c=2π
Bohachevsky function	$f(x) = x_i^2 + 2x_i^2 - 0.3\cos(3\pi x_1) - 0.4\cos(3\pi x_2) + 0.7$	$x_i \in [-10, 10]$ for ∀ i =1, 2
Mccormick function	$f(x) = \sin(x_1 + x_2) + (x_1 - x_2)^2 - 1.5x_1 + 2.5x_2 + 1$	$x_i \in [-10, 10]$ for ∀ i =1, 2

3.2.7.2 Results

A separate practice is performed for each optimization function and also produced a separate outcome to check the performance individually. Once the community partitioning operation is completed on both the dataset, it is observed that the synthesized and real datasets have a total of 5 and 24 communities, respectively. Next, the proposed social network-based whale optimization algorithm is performed on both the datasets and measured the opinion leaders in each community separately and globally. Further, ranked the top-10 opinion leaders having the maximum reputation based on ten benchmark optimization functions and standard centrality measures after normalization on both the datasets, respectively, as shown in Table 3.10 and Table 3.11.

In both the Tables, each user's global minima is calculated using a standard benchmark optimization function. The global minima are used for calculating the reputation of the user. Here, only top-10 opinion leaders are discussed with their several centrality measures and reputation obtained. In this analysis, it is worth mentioning to state which techniques are better or worst for identifying the opinion leader. For example, in column 2 of Table 3.10, node #62 having ranked one with a reputation of 0.45548 for the synthesized dataset; So, node #62 is the top opinion leader according to the 'Easom' function. Similarly, in column Table 3.11, node #4482 has a reputation of 0.87981 and secured the real dataset's top position.

Table 3.10. Top-10 opinion leaders with the highest reputation r using standard benchmark optimization functions and standard centrality measures on the synthesized dataset

Rank	Node id	Easom	Node id	Ackley	Node id	Matyas	Node id	Bukin	Node id	Beale	Node id	Dixon price	Node id	Three hump camel
1	62	0.45548	23	0.21156	49	0.44322	10	0.2213	24	0.66254	86	0.52014	25	0.62541
2	72	0.45547	58	0.21155	83	0.44321	45	0.22129	74	0.66253	59	0.52013	46	0.6254
3	28	0.45547	47	0.21154	15	0.4432	75	0.22128	48	0.66252	43	0.52012	34	0.62539
4	74	0.45546	62	0.21153	93	0.44319	57	0.22127	42	0.66251	6	0.52011	85	0.62538

5	**48**	**0.45544**	**29**	**0.21152**	**80**	**0.44319**	**68**	**0.22126**	**26**	**0.6625**	**42**	**0.5201**	**62**	**0.62537**
6	**13**	**0.45544**	**53**	**0.21152**	**61**	**0.44317**	**73**	**0.22125**	**66**	**0.66249**	**10**	**0.52009**	**18**	**0.62536**
7	**70**	**0.45543**	**41**	**0.2115**	**44**	**0.44316**	**98**	**0.22124**	**83**	**0.66248**	**8**	**0.52008**	**37**	**0.62535**
8	**42**	**0.45542**	**43**	**0.21149**	**85**	**0.44315**	**22**	**0.22123**	**37**	**0.66247**	**37**	**0.52007**	**36**	**0.62534**
9	**34**	**0.45541**	**21**	**0.21148**	**9**	**0.44313**	**39**	**0.22122**	**24**	**0.66246**	**16**	**0.52006**	**10**	**0.62533**
10	**81**	**0.455**	**4**	**0.21148**	**131**	**0.44313**	**44**	**0.22121**	**11**	**0.66245**	**77**	**0.52005**	**8**	**0.62532**

Rank	**Node id**	**Drop wave**	**Node id**	**Bohachevsky**	**Node id**	**Mccormick**	**Node id**	**BC**	**Node id**	**DC**	**Node id**	**CC**	**Node id**	**EV**	**Node id**	**PR**
1	**74**	**0.71025**	**99**	**0.42213**	**89**	**0.82139**	62	0.41244	74	0.69521	74	0.71452	19	0.38547	74	0.74235
2	**72**	**0.71024**	**10**	**0.42212**	**12**	**0.82138**	44	0.41243	25	0.6952	28	0.71451	91	0.38546	83	0.74234
3	**57**	**0.71023**	**74**	**0.42211**	**35**	**0.82137**	7	0.41242	18	0.69519	25	0.7145	12	0.38545	32	0.74233
4	**62**	**0.71022**	**91**	**0.4221**	**6**	**0.82136**	28	0.41241	10	0.69518	18	0.71449	74	0.38544	18	0.74233
5	**29**	**0.71022**	**28**	**0.42209**	**10**	**0.82136**	25	0.4124	28	0.69517	10	0.71448	15	0.38543	43	0.74231
6	**47**	**0.7102**	**33**	**0.42208**	**12**	**0.82134**	10	0.41239	25	0.69516	50	0.71447	46	0.38542	10	0.7423
7	**29**	**0.71019**	**21**	**0.42207**	**28**	**0.82133**	34	0.41238	25	0.69516	11	0.71446	13	0.38542	28	0.74229
8	**41**	**0.71018**	**7**	**0.42206**	**2**	**0.82132**	78	0.41237	86	0.69514	12	0.71445	6	0.3854	25	0.74228
9	**85**	**0.71018**	**62**	**0.42205**	**19**	**0.82132**	74	0.41235	2	0.69513	95	0.71444	26	0.38539	12	0.74228
10	**82**	**0.71016**	**90**	**0.42204**	**69**	**0.8213**	12	0.41235	53	0.69513	25	0.71443	94	0.38539	86	0.74226

Table 3.11. Top-10 opinion leaders with the highest reputation r using standard benchmark optimization functions and standard centrality measures on the real dataset

Rank	**Node id**	**Easom**	**Node id**	**Ackley**	**Node id**	**Matyas**	**Node id**	**Bukin**	**Node id**	**Beale**	**Node id**	**Dixon price**	**Node id**	**Three hump camel**
1	**4482**	**0.87981**	**4482**	**0.52352**	**4482**	**0.74251**	**4482**	**0.68523**	**4482**	**0.61138**	**4482**	**0.82143**	**4482**	**0.64582**

2	1869	0.87980	1308	0.52347	2123	0.74249	2123	0.68521	763	0.61125	2123	0.82140	2123	0.64581
3	1308	0.87978	2123	0.52344	1308	0.74245	1308	0.68519	2123	0.61118	1308	0.82136	1308	0.64577
4	3389	0.87975	1869	0.52342	1869	0.74240	2123	0.68518	1308	0.61108	1869	0.82135	1869	0.64576
5	2123	0.87974	763	0.52294	763	0.74238	763	0.68515	6713	0.61107	2123	0.82131	763	0.64574
6	763	0.87972	3389	0.52291	3389	0.74235	3389	0.68510	5302	0.61101	763	0.82129	3389	0.64571
7	2990	0.87971	2990	0.52289	2290	0.74232	2290	0.68509	75	0.61092	2290	0.82126	2290	0.64570
8	6713	0.87969	6713	0.52284	6713	0.74229	6713	0.68507	3389	0.61089	6713	0.82121	6713	0.64567
9	5302	0.87965	75	0.52282	5302	0.74221	75	0.68503	1869	0.61087	5302	0.82119	3110	0.64561
10	75	0.87961	3110	0.52279	3110	0.74216	3110	0.68501	2990	0.61086	3110	0.82116	5302	0.64558

Rank	Node id	Dropwave	Node id	Bohachevsky	Node id	Mccormick	Node id	BC	Node id	DC	Node id	CC	Node id	EV	Node id	PR
1	4482	0.71224	4482	0.55894	4482	0.71003	4482	0.57	4482	0.77255	4482	0.40264	4482	0.6366	4482	0.37798
2	1308	0.71223	1308	0.55893	1308	0.70998	2123	0.56999	2123	0.77253	763	0.40263	2123	0.6366	1869	0.37796
3	2123	0.71220	2123	0.55891	2123	0.70997	1308	0.56998	1308	0.77251	2123	0.40263	1308	0.6366	1308	0.37795
4	1869	0.71217	1869	0.55890	763	0.70995	1869	0.56997	1869	0.77251	1308	0.40262	75	0.6365	3389	0.37794
5	763	0.71215	763	0.55889	1869	0.70993	763	0.56997	763	0.7725	6713	0.40261	763	0.6365	2123	0.37793
6	3389	0.71211	3389	0.55888	2990	0.70991	3389	0.56996	3389	0.77249	5302	0.40261	3389	0.6365	763	0.37792
7	2990	0.71209	2990	0.55886	6713	0.70990	2290	0.56995	2290	0.77249	3389	0.4026	2990	0.6365	2990	0.37791
8	6713	0.71208	6713	0.55885	75	0.70988	6713	0.56995	6713	0.77248	75	0.40259	6713	0.6365	6713	0.37791
9	5302	0.71205	75	0.55883	5302	0.70985	5302	0.56994	3110	0.77247	1869	0.40259	1869	0.6365	5302	0.3779
10	3110	0.71204	3110	0.55882	3110	0.70984	3110	0.56993	5302	0.77246	4910	0.40257	3332	0.6365	75	0.37789

Next, the top opinion leaders are assessed with each community's maximum reputation on a synthesized and real dataset. As discussed earlier, that synthesized and real dataset has 5 and 24 communities, respectively. Again, the ten benchmark optimization functions are used to measure the reputation and also calculated various centrality measures for each user in every community. After calculating the normalized reputation and centrality measure, the top opinion leaders in each community for the synthesized and real dataset are shown in Table 3.12 and Table 3.13, respectively. For example, in Table 3.12, node #72 is considered the top opinion leader for community #2 based on the ‘Easom’ optimization function, i.e., node #72 having a higher reputation synthesized dataset. Similarly, in Table 3.13, node #321 is considered a top opinion leader in community #1 based on the ‘Easom’ optimization function.

Table 3.12. A top opinion leader in each community having the highest reputation r using standard benchmark optimization functions and centralities measures on the synthesized dataset

Community	Easom	Ackley	Matyas	Bukin	Beale	Dixon price	Three hump camel	Drop wave	Bohachvsky	Mccormick	BC	DC	CC	EV	PR
1	**62**	**23**	**93**	**68**	**48**	**86**	**125**	**74**	**91**	**10**	62	74	74	15	32
2	**72**	**58**	**80**	**10**	**14**	**43**	**46**	**72**	**12**	**89**	44	25	28	19	74
3	**28**	**47**	**49**	**45**	**26**	**43**	**34**	**57**	**99**	**12**	7	18	25	91	9
4	**74**	**62**	**83**	**75**	**24**	**6**	**85**	**62**	**10**	**35**	28	10	18	12	18
5	**48**	**29**	**15**	**57**	**74**	**42**	**62**	**12**	**74**	**6**	25	28	5	74	11

Table 3.13. A top opinion leader in each community having the highest reputation r using standard benchmark optimization functions and centralities measures on the real dataset

Community	Easom	Ackley	Matyas	Bukin	Beale	Dixon price	Three hump camel	Drop wave	Bohachevsky	Mccormick	BC	DC	CC	EV	PR
1	**321**	**321**	**1089**	**765**	**923**	**1089**	**3422**	**3098**	**2286**	**275**	745	321	3066	2331	1089
2	**3389**	**3389**	**3389**	**3389**	**3389**	**2005**	**3389**	**3389**	**3389**	**3389**	3389	5342	3389	3276	118

3	**165**	**3033**	**165**	**659**	**1004**	**165**	**2238**	**738**	**3033**	**472**	6345	165	165	3033	3033
4	**6713**	**6713**	**6713**	**6713**	**6713**	**6713**	**6713**	**6713**	**6713**	**6713**	6713	6713	6713	6713	6713
5	**5983**	**683**	**592**	**3875**	**2285**	**186**	**24**	**186**	**683**	**5983**	24	24	186	1522	24
6	**1522**	**2640**	**72**	**1522**	**1522**	**72**	**2640**	**72**	**2640**	**744**	72	1522	72	72	72
7	**3110**	**3110**	**3110**	**3110**	**3110**	**3110**	**4434**	**3110**	**3110**	**3110**	3110	4434	4434	3110	3110
8	**75**	**75**	**694**	**75**	**75**	**1693**	**2529**	**75**	**75**	**75**	694	75	75	694	377
9	**2268**	**4430**	**2268**	**2268**	**1975**	**2268**	**4430**	**1975**	**1975**	**2268**	4430	2268	4430	2268	2268
10	**1869**	**1869**	**1869**	**3358**	**1869**	**1869**	**1869**	**1869**	**1869**	**1869**	3358	1869	1869	1869	1869
11	**763**	**763**	**763**	**763**	**763**	**763**	**763**	**763**	**763**	**763**	763	763	763	763	763
12	**1373**	**1373**	**1373**	**1373**	**1373**	**3360**	**1373**	**1373**	**173**	**1373**	1373	1373	1373	1373	1373
13	**2069**	**2069**	**2069**	**2069**	**522**	**522**	**2069**	**522**	**522**	**2069**	522	522	522	2069	2069
14	**4482**	**4482**	**4482**	**4482**	**4482**	**4482**	**4482**	**4482**	**4482**	**4482**	4482	4482	4482	4482	4482
15	**3398**	**3398**	**3398**	**3398**	**368**	**368**	**368**	**3398**	**368**	**3398**	368	3398	368	853	3398
16	**3968**	**725**	**1196**	**1196**	**2784**	**1196**	**2784**	**3968**	**2784**	**3968**	2784	1196	2784	3968	992
17	**2990**	**2990**	**2290**	**2290**	**2990**	**2290**	**3318**	**2290**	**2290**	**2990**	2990	3318	2990	2990	2990
18	**1308**	**1308**	**1308**	**1308**	**1308**	**1308**	**1308**	**1308**	**1308**	**1308**	1308	1308	1308	1308	779
19	**5589**	**6439**	**5589**	**6439**	**6439**	**6439**	**6439**	**5589**	**5589**	**6439**	5589	6439	6439	6439	6439
20	**149**	**2004**	**149**	**149**	**5593**	**149**	**149**	**1905**	**149**	**149**	149	149	149	2004	2004
21	**5302**	**5302**	**5302**	**666**	**5302**	**5302**	**5302**	**5302**	**666**	**5302**	5302	5302	5302	666	666
22	**3029**	**4729**	**3889**	**2748**	**4729**	**4729**	**4729**	**3089**	**39**	**3869**	1984	1578	6823	387	2748
23	**2123**	**2123**	**2123**	**2123**	**2123**	**2123**	**2123**	**2123**	**2123**	**2123**	2123	2123	2123	2123	2123
24	**47**	**47**	**338**	**47**	**338**	**5782**	**5782**	**47**	**5782**	**5782**	47	3372	47	5731	47

The execution of the whale algorithm optimization process is represented using different optimization functions. The graphical illustration of the global minima test function, often known as the optimization function, is used during the social network-based whale optimization algorithm, as shown in Fig. 3.14. Each function is a continuous unimodal or multimodal function represented in 2-D space. A unimodal test function has only one mode, while a multimodal test function has more than one mode. Each graph of Fig. 3.14(a-j) captures the global minima in the subspace-based on its mathematical expression, and the x-axes indicate the input range in the interval [-10, 10], and y-axes indicate the value of f(x). Therefore, each graph presents a specific outcome viewpoint for the social network-based whale optimization algorithm.

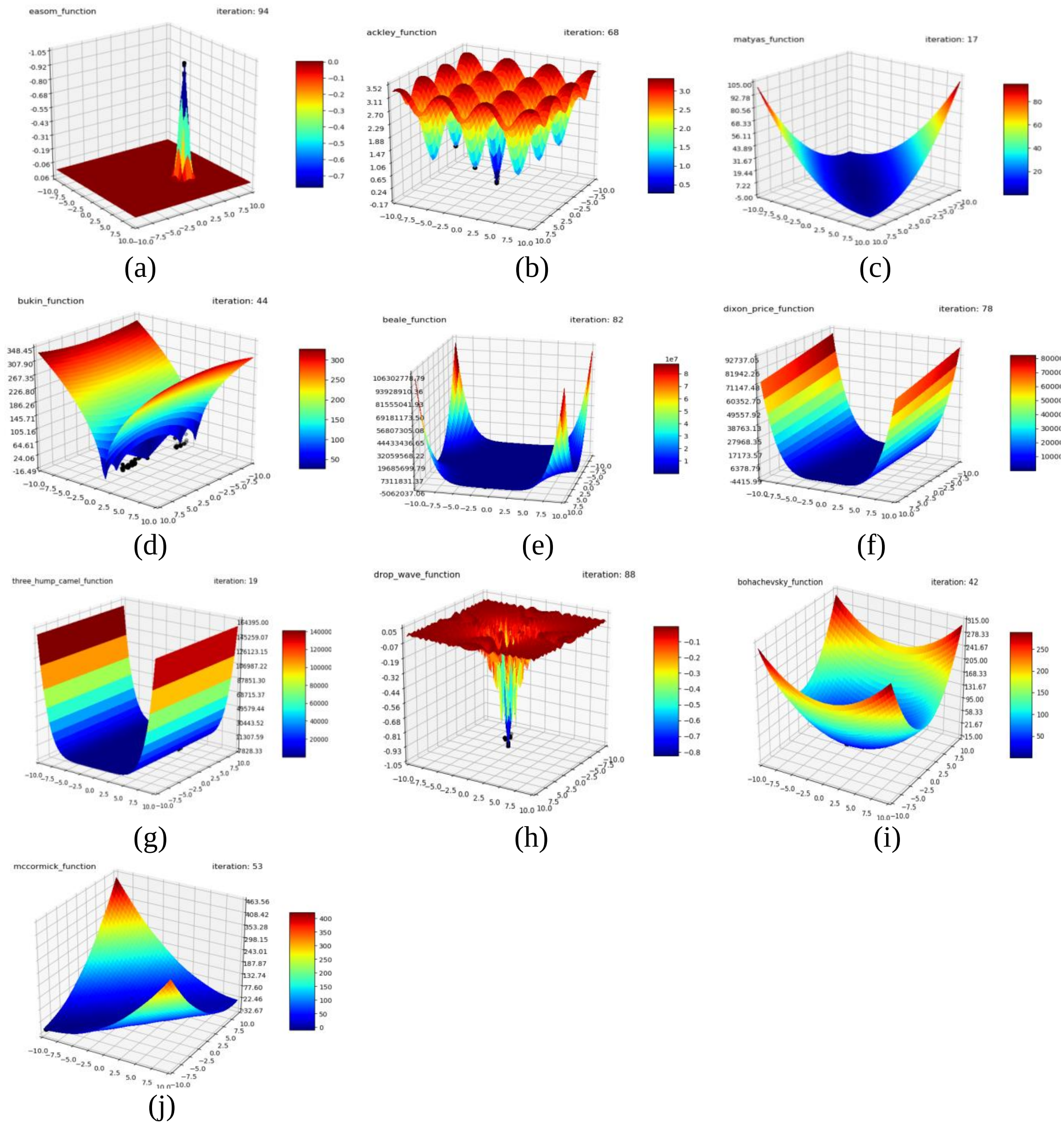


Fig. 3.14. Whale optimization algorithm execution using (a) Easom function, (b) Ackley function, (c) Matyas function, (d) Bukin function, (e) Beale function, (f) Dixon price function, (g) three hump camel function, (h) Drop wave function, (i) Bohachevsky function, (j) Mccormick function

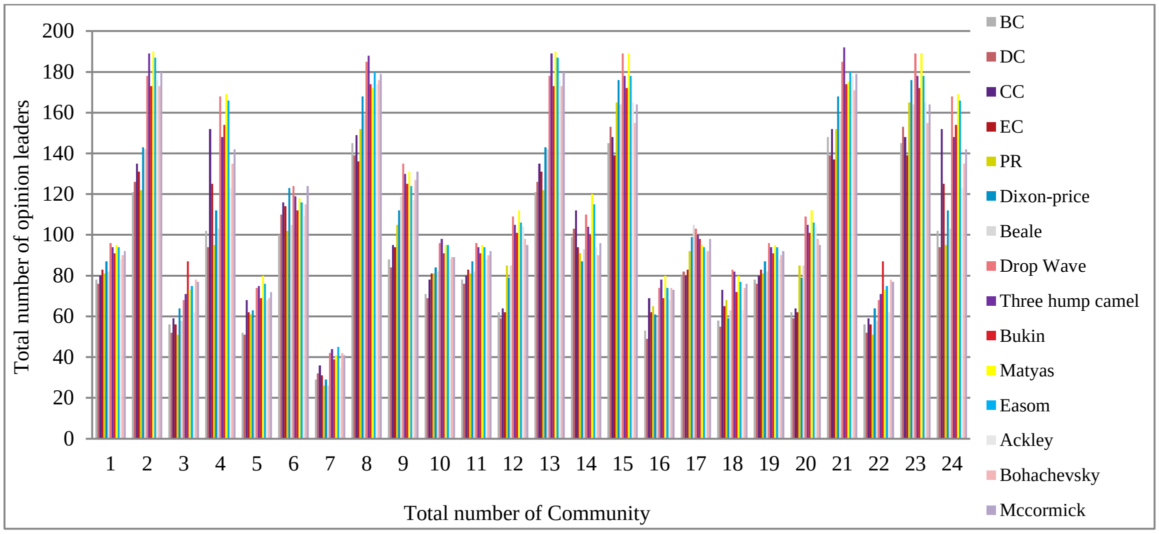


Fig. 3.15. Total number of opinion leader identified by standard benchmark functions and SNA measures in each community in the real dataset

Furthermore, it is identified that the total number of opinion leaders varies in each community measured by standard benchmark optimization functions and SNA measures, as shown in Fig. 3.15. It is observed that the social network based whale optimization found out the more users who have a high reputation in the network. For the real dataset, ‘Matyas,’ ‘Three hump camel,’ and ‘Drop wave’ function produced better results and identified the highest number of opinion leaders in most communities.

In this experiment, It is also investigated that as the control parameter's value changed, the total number of opinion leaders identified by each optimization method is too assorted. Fig. 3.16 shows the total number of adopted opinion leaders by each optimization methods defined as the value of the control parameter varies from 0 to 2. It is experiential that the value of the control parameter depends on the total number of iterations. If the whole number of repetitions is too high, the control parameter's value is also too high. It is also pointed out that if the control parameter value is one, it identified the maximum number of opinion leaders for each optimization function, i.e., a synchronized balance needed between the total number of iterations and control parameter. For example, when the value of the control parameter is one, ‘Matyas,’ ‘Drop wave,’ and ‘Easom’ are the top-3 functions that identified the maximum number of opinion leaders. In contrast 'Beale' function identified the minimum number of opinion leaders in the real dataset.

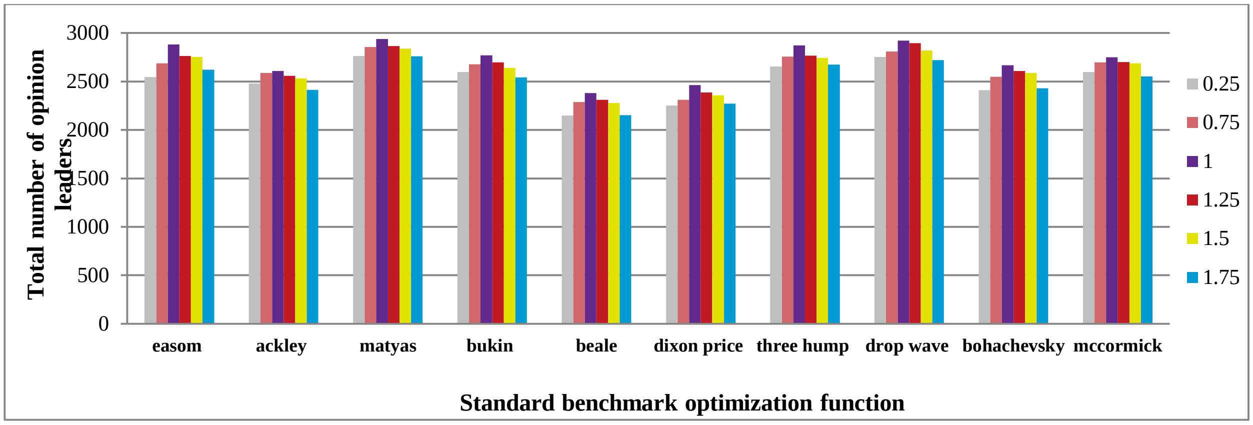
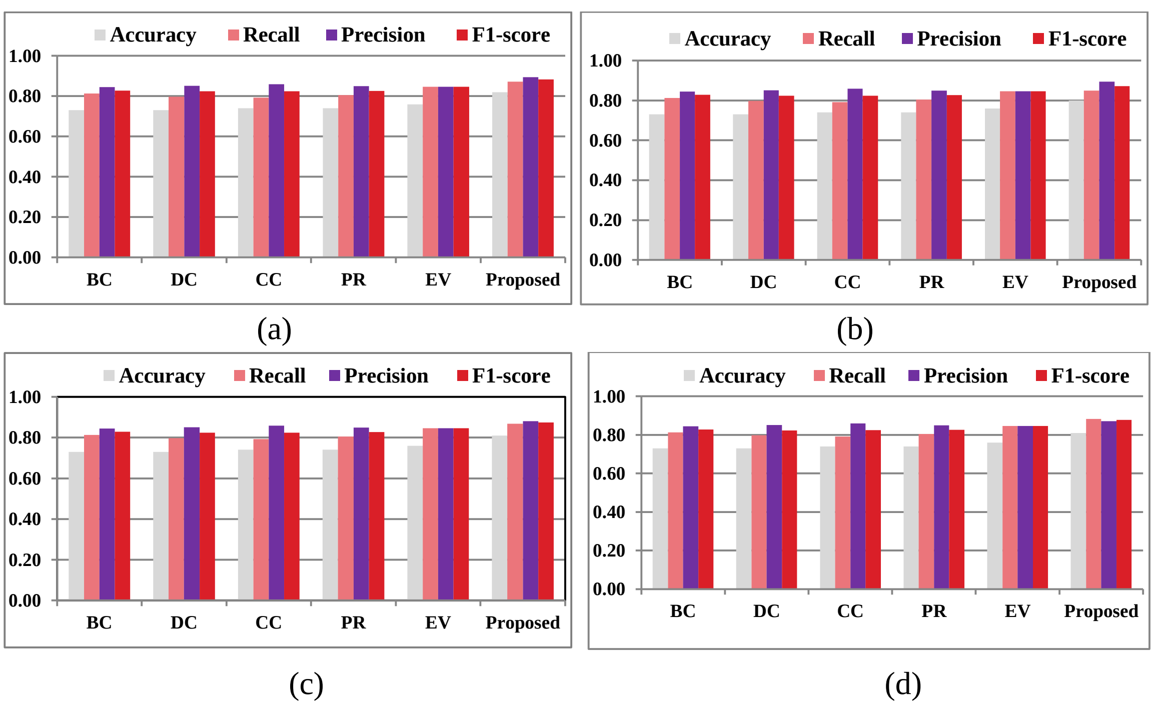


Fig. 3.16. Total number of opinion leaders identified by different standard benchmark functions as the control parameter a, changes in real dataset

For authorizing the validity of the algorithm, the results are compared with the other SNA measures used for the same dataset.



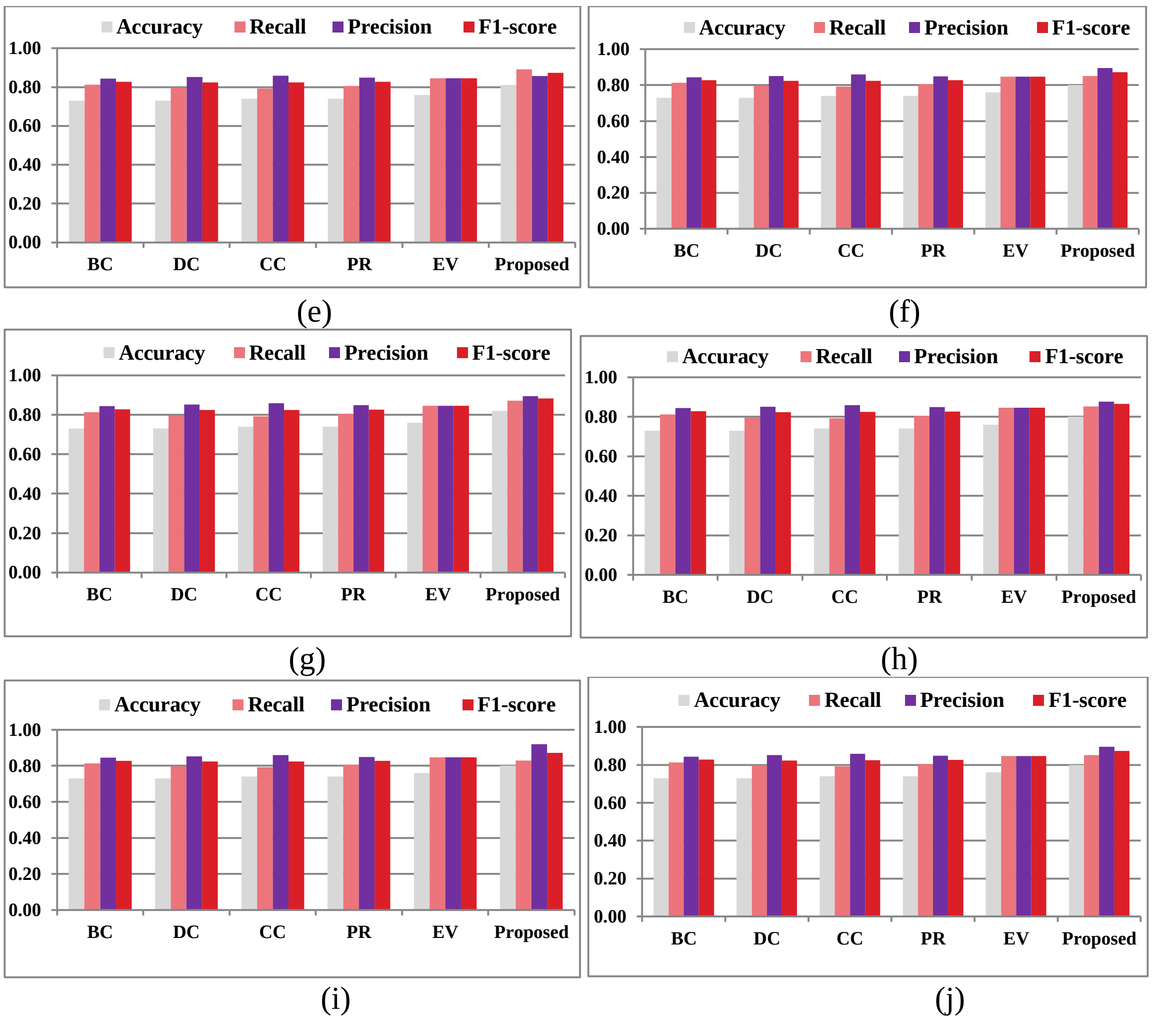


Fig. 3.17. Comparison of accuracy, recall, precision, and F_1-score with different standard benchmark (a) Easom function, (b) Ackley function, (c) Matyas function, (d) Bukin function, (e) Beale function, (f) Dixon price function, (g) three hump camel function, (h) Drop wave function, (i) Bohachevsky function, (j) Mccormick function

Sometimes, it may be possible that the traditional SNA measures provide better results as compared to advance approaches, depending upon the type of network structure. Also, the implementations of SNA measures also relatively straightforward and understandable, but, empirically, it is observed that as the size of the dataset grows gradually, these measures take too much execution time. In the case of the real dataset, the size of the dataset is too large, and it is detected that the proposed approach is better than the other SNA measures. For example, in Fig. 3.17(a), six different bar clusters has made; the first bar cluster for betweenness centrality, second bar cluster for degree centrality, third bar cluster for closeness centrality, fourth bar cluster for page rank, fifth bar cluster for eigenvector centrality, and sixth bar cluster for a proposed approach using Easom optimization function.

Similarly, in each part, the same procedure applied in which the sixth bar cluster indicates the proposed approach's outcome value using different optimization functions. It is evaluated that the proposed approach is better in every part. It is also observed that these optimization functions might not always produce better results for all types of datasets because network structure, user's characteristics, types of relationships, and network dynamics also change in the various networks. From the above result, it is inferred that

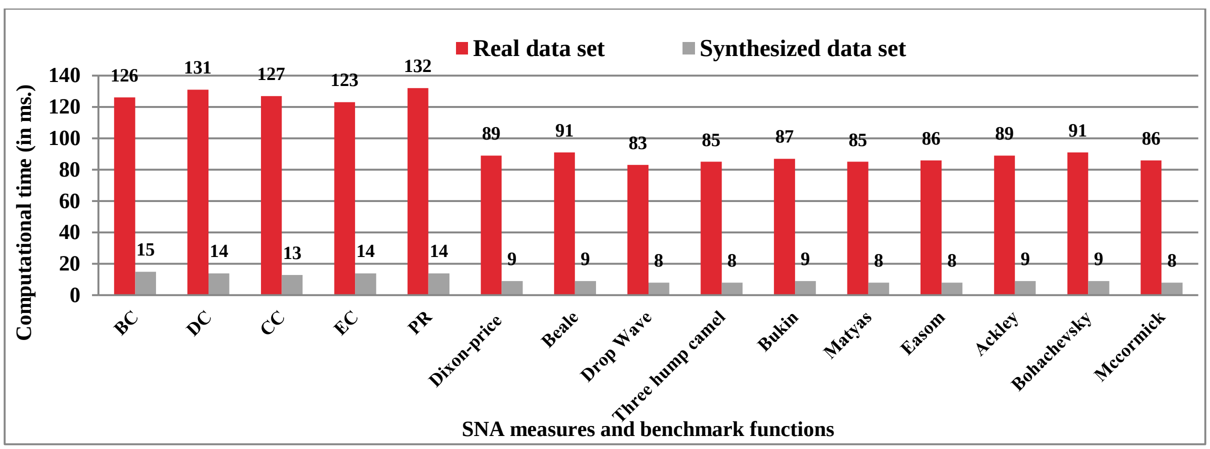


Fig. 3.18. Comparison of computational time with different standard benchmark optimization function on the real and synthesized datasets

the proposed algorithm is better than other SNA measures in terms of accuracy, precision, recall, and F1-score using all the optimization techniques that are used in the whale optimization algorithm. In this study, it is also analyzed that the total computational time needed by the proposed algorithm for executing the entire process in both the datasets is also very less as compared to standard SNA measures, as shown in Fig. 3.18.

Table 3.14. Comparison between proposed and other community detection methods for real and synthesized dataset

Parameters	**Louvain**		**InfoMap**		**FastGreedy**		**Walktrap**		**Makov cluster**		**Proposed**	
	Real	**Synthesized**	**Real**	**Synthesized**	**Real**	**Synthesized**	**Real**	**Synthesized**	**Real**	**Synthesized**	**Real**	**Synthesized**
Total number of nodes	7115	100	7115	100	7115	100	7115	100	7115	100	**7115**	**100**
Modularity	0.424	0.211	0.455	0.265	0.402	0.115	0.395	0.112	0.435	0.226	**0.398**	**0.187**
Node attribute similarity	0.618	0.264	0.625	0.258	0.585	0.289	0.652	0.274	0.624	0.281	**0.518**	**0.255**
Common neighbor similarity	0.775	0.472	0.712	0.452	0.769	0.483	0.721	0.466	0.684	0.479	**0.662**	**0.459**
Total running time (in sec.)	538	16.2	520	14.5	526	14.7	548	17.3	563	18.8	**506**	**12.7**
Average community density	0.856	0.526	0.956	0.556	0.758	0.493	0.834	0.519	0.795	0.512	**0.699**	**0.471**

3.2.8 Performance of social network based whale optimization algorithm

The Social Network based Whale Optimization Algorithm (SNWOA) has stronger steadiness, a quicker rate of convergence, and more accuracy as compared to other nature-inspired meta-heuristic algorithms used for identifying the opinion leaders. In the previous work, the opinion leaders are determined using the firefly algorithm in each community locally and in the social network globally. Each algorithm has its features and dynamics to solve a particular real-world problem, and no algorithm is appropriate for solving all issues. So, both the algorithms are analyzed based on diffusion rate, convergence rate, accuracy, micro-average precision, and macro-average precision. The diffusion rate determines how quickly the information spreads in the network over a while. The convergence rate determines the speed of an algorithm, i.e., the total number of iterations needed to reach the prescribed limit. If the rate of convergence is higher, fewer iterations are required for obtaining the outcome. Accuracy determines how many users in the dataset correctly classify as opinion leaders based on network ground-truth. In Micro-average precision, the individual true positive, false positive, and false negative are added to the network for different groups and then perform the mathematical operation for measurement. Now, both the algorithms are compared based on the mentioned parameters in Table 3.15. In this table, it is analyzed that the whale optimization algorithm offered better results as compared to the firefly algorithm for both real and synthesized datasets.

Table 3.15. Comparison of whale optimization algorithm with a firefly optimization algorithm for real and synthesized dataset

Characteristics	**Whale optimization**		**Firefly optimization**	
	Real dataset	**Synthesized dataset**	**Real dataset**	**Synthesized dataset**
Diffusion rate	**0.458902**	**0.689038**	0.397801	0.667945
Rate of convergence	**5.008622**	**7.075989**	4.654073	6.080763
Accuracy	**0.862859**	**0.830117**	0.790865	0.775116
Micro-average precision	**0.920762**	**0.845836**	0.860782	0.790443
Macro-average precision	**0.918915**	**0.820814**	0.849801	0.770679

The average error value achieved by both the algorithms is also compared for each optimization functions. The average error value is the average of all the deviations between actual and measured optimum value delivered in each iteration. The main reason behind this error is the initial value of the coefficients and parameters that are chosen for the function. If the initial value of the coefficients and parameters are too high, it may lead to distance from the optimal solution. The size of the network also matters for finding the optimal solution in the case of metaheuristic. Table 3.16 discovered that the 'Matyas' optimization function offered better results, i.e., produced lower average error value while the 'Dixon-price' function offered worst results, i.e., produced higher average errors value for the whale optimization algorithm. In the firefly algorithm for the same dataset, the 'Three hump camel' and 'Drop wave' function provided the best and worst results, respectively.

Table 3.16. Averaged errors value achieved by whale and firefly algorithm for each optimization functions

Function	**Whale optimization**		**Firefly optimization**	
	Real dataset	**Synthesized dataset**	**Real dataset**	**Synthesized dataset**
Dixon-price	**0.086769**	**0.099761**	0.112519	0.132512
Beale	0.070056	0.081583	0.115628	0.135262
Drop Wave	0.076858	0.082035	**0.116253**	**0.135541**
Three hump camel	0.075164	0.080089	**0.105843**	**0.121514**
Bukin	0.078902	0.087005	0.109347	0.130054
Matyas	**0.074678**	**0.079982**	0.107774	0.123995
Easom	0.078766	0.085021	0.110025	0.124557
Ackley	0.080067	0.092102	0.108432	0.128947
Bohachevsky	0.081005	0.094580	0.110584	0.133802
Mccormick	0.081740	0.093958	0.115444	0.134440

Further, both algorithms' average running time for each optimization function is also compared, as shown in Table 3.17. In this table, it is found that all the optimization functions except 'Beale' and 'Bukin' consumed less time for the whale optimization algorithm as compared to the firefly algorithm. Therefore, in a nutshell, it is inferred that the proposed social network-based whale optimization algorithm is better than the firefly algorithm in almost all the features.

Table 3.17. Averaged running time (in sec.) attained by whale and firefly algorithm for each optimization functions

Function	Whale optimization		Firefly optimization	
	Real dataset	Synthesized dataset	Real dataset	Synthesized dataset
Dixon-price	**89**	**9**	93	10
Beale	91	9	**87**	**8**
Drop Wave	**83**	**8**	88	9
Three hump camel	**85**	**8**	91	8
Bukin	87	**9**	**85**	9
Matyas	**85**	**8**	88	9
Easom	**86**	**8**	90	9
Ackley	**89**	**9**	92	9
Bohachevsky	**91**	**9**	97	10
Mccormick	**86**	8	91	9

3.3 Chapter summary

In this chapter, two nature-inspired algorithms are addressed to find the practical and optimal opinion leaders in the different online social networks. In the firefly algorithm, initially, the communities are identified using the modified Louvain community partitioning algorithms in which the concept of clustering coefficient is associated to find out the communities. Next, the local and global opinion leaders are discovered using the firefly algorithm that produces a better result than other SNA measures. The result indicates that the algorithm finds the optimal opinion leader. Further, a new social network based nature-inspired whale optimization algorithms explained with different standard benchmark optimization functions to identify the top-N opinion leaders in the social network. Initially, the objective function for each user using their distance and centrality is measured. Next, the community partitioning algorithm is implemented to determine the communities in the datasets. Further, the whale optimization technique applied with different optimization to estimate the top-10 opinion leader at the local and global stage. The proposed algorithms' main asset is that as the number of users increases, the algorithm's accuracy and efficiency also increase because more information about the other user's vector position is accumulated.

There are lots of further dimensions suggested by both approaches. The first one is exploring the other nature-inspired algorithm [159], [160] that might yield better performance and accuracy. Another one is to investigate users' other unique characteristics such as social status, culture, financial background, technical knowledge, experience, trust value [161], relationship, global identity, and many more to find more precise and concise results for online social networks.

Chapter 4

Recognition of opinion leader's coalitions for information diffusion: Game theory approach

Game theory provides more irrefutable and concise information about human decisions using multiple optimal strategies. This chapter dealt with the opinion leader detection dilemma in the online social network with the evolutionary game theory approach. Section 4.1 reveals the overview of the entire chapter. Section 4.2 gives details of game theory and a brief of Shapley value. Section 4.3 provides the details of a novel Game theory-based Opinion Leader Detection (GOLD) algorithm to identify users with the maximum synergy declared as the coalition of opinion leaders. All the inventive and distinctive solution is defined to measures the individual payoff using the distance-based centrality parameter. Section 4.4 explains the calculation of Shapley value for each user to identify the maximum marginal contribution and determine each coalition's maximum synergy. The complexity of the entire procedure is illustrated in section 4.5. Section 4.6 demonstrates the description of the datasets along with their implementation outcomes. Finally, section 4.7 concludes the summary of the entire chapter.

4.1 Overview

Game theory played a significant role in finding out the opinion leader in the social network. The action performed by one user affects the outcomes of another user that they intended to achieve. A game theory approach is logical learning of the strategic interaction of the coherent user. In the real-life, the users do not know whether their actions affect the other user's outcomes, but they must understand that they know their efforts. The game theory's primary assumption is that the user must be rational and attempted to maximize their payoff. The social network also adopted the game theory approach, in the same way, to find the opinion leader in promoting and capitalizing on the growth of products in the real-world commercial market [162], [163]. Game theory also approaches beneficial to identify the different types of power centrality and relationships between the users in the social network [164], [165]. Various issues related to network privacy and security are also resolved by game theory approaches [166].

For instance, consider a network with a total of 16 nodes, as shown in Fig.4.1. The top nodes with a higher degree centrality are $v_5, v_7, v_8, v_{11},$ and v_{14}. Here, the node v_2 with higher betweenness centrality and node $v_{11}$with higher closeness centrality, but it is not the right approach to analyze the node's importance based on a single attribute. Only one or two nodes are not responsible for the information diffusion, and a set of nodes disseminate the information in the network.

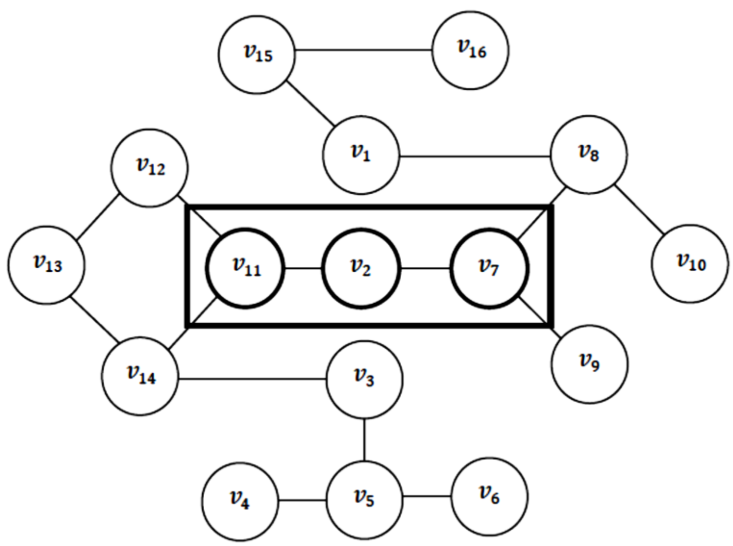


Fig. 4.1. Representation of synergetic coalition of nodes

It is observed that the synergetic coalition of nodes (v_2,v_7, v_{11}) could spread or stop the information dissemination depending upon circumstances. If any of the false information is flooded by any other node, the coalition of these nodes can easily shatter information propagation.

4.2 Introduction of game theory

Game theory offers a prescribed, structured, organized, analyzed, and systematically evaluated scenario that includes some players and a set of rules followed by each player [167]. Players play the game according to the rules. Whenever the player chooses a strategy and performs an action, an outcome produces. There is a payoff associated with each strategy stroke and can be of the monetary or non-monetary type, i.e., in the form of happiness, joy, or completeness. There are also some predetermined rules to play a game, but either the player may change its plan, or the environment and conditions may change results insufficient desired payoff. An approach is called optimum if it increases the player's payoff in each move. So, some strategies move are exquisite and produce higher payoff. Game theory also follows the no-cheating phenomenon, i.e., no player can deceive other players in the game. So, it depicts the strategies that show how a player can win the game without any cheating [168], [169]. Game theory classified the game into two categories: **Normal form game and Extensive form game.**

4.2.1 Normal form game

There are a finite set of players in a normal form game, and each player has a limited set of strategies. A payoff function assigns some payoff to each player depending on the player's technique and another player's strategy. For example, in the Prisoner's Dilemma, two persons with their plans cooperate(C) and defeat (D). If any of the people change their policy, both lose their optimal payoff [170].

4.2.2 Extensive form game

An extensive form game is a well-formed tree is defined in which every player has information about the other players like their strategies, possible outcomes of a sequence of moves, payoff associated with all game results. The tree structure indicates the graphical representation of a player's decision at a specific moment. The tree structure includes a root node that represents nature to move and responsible for the starting of the game. Each leaf node represents the n-tuple payoff vector that consists of the payoff associated with each player obtained at the end of the moves. Each node of the tree indicates the possible level of the game from which the player chooses its strategy for achieving the next level. Each edge of the tree depicts the possible action taken by the player at each level. Once a user reaches a particular level, the player uses a probability distribution function over an edge to get another stage. The tree's terminal nodes contain the final payoff associated with every player involved in the game and exemplify the game's ending [171], [172].

For example, two players A and B, play with two strategies X and Y, in Fig. 4.2. At the ending of the game, both players receive some payoff. If player A chooses strategy Y and player B adopts strategy X, the yield might be one for player A and two for player B. Similarly, if both the payer chooses the strategy Y, both get the payoff zero.

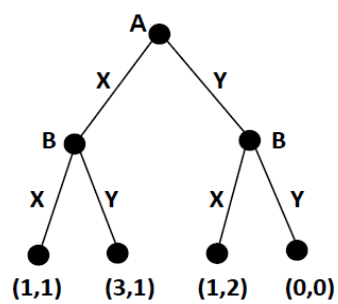


Fig. 4.2. The trustor-trustee tree structure for extensive form game

4.2.3 Shapley value

Trust is a prominent feature that affects the degree of relationship and makes a strong or weak connection between the users. For calculating the trust among the users, the concept of Shapley value is used in the research [173]. The Shapley value measures the expected payoff obtained by the player in the coalition game theory. The Shapley value represents the contribution of the player in a coalition having n number of players [174]. In other words, it is defined as that if the group of people participated in a coalition and got some payoff, how the payoff divided among the players fairly. Shapley's main advantage is to provide the expected marginal payoff based on the participation of the player in a game

[175], [176]. There are various ways to calculate the Shapley value with some constraints [177]. In this examination, a hypothesis is defined as whether all the players belong to the same coalition or are nearby to at-least x-neighbors in the coalition. The main reason behind this assumption is that in the real world also, users generally interact only with those persons who follow the same behavior as by them and likely to make a group with them. In this case, a game can be represented as (N, v) where N represents the total number of players in the game, i.e., N = {1, 2, 3, ..., n} while v indicates the function $v \in R^{2^n-1}$ served using Eq. (4.1).

$$v = \begin{Bmatrix} 0 & \text{if } C = 0 \\ \{v \in C \mid |N(v) \cap C| \geq x\} & \text{if } C > 0 \end{Bmatrix} \quad \ldots (4.1)$$

It is evident that if deg(node) < x, all the node belongs to the same coalition otherwise divide the network into other coalitions. So this is required that the deg(node) ≥ (x+1) for more precise calculation. In the coalition game, every player has its valuable reserve, and once the player alliances in the game, it may produce some collective synergy that is higher or may be lower than the sum of their resource at time t. For example, if two players I and j have an individual payoff in a game, it is 3 and 2, respectively. Still, if both players collate in the game, the collective synergy might be six (multiplicative factor), which is higher than the sum of their payoff [178]. Therefore, Shapley value analyzed the contribution of each user once they come together and make a cluster. The Shapley value of an individual based on payoff can be represented using Eq. (4.2).

$$SP(v) = \sum_{C\epsilon 1}^{N-i} \frac{|C|!(n-|C|-1)!}{n!} \{v(C \cup i) - v(C)\} \quad \ldots (4.2)$$

In the above equation $\{v(C \cup i) - v(C)\}$ represents the marginal payoff received by the user in a collation game *C*. Hence, the above equation depicts weight to the expected contribution.

4.3 Proposed methodology

The game theory approach is categorized into two forms; extensive and normal form, respectively. The extensive game theory form is used to generate a trustor-trustee tree based on trust for implementing the proposed strategy. The trust and centrality measures are used to define the user characteristics in the network [179]. It is hypothesized that each user behaves like a player in the network. Trust and other centrality measures are considered attributes that help to compute the marginal contribution in the game.

Initially, the degree of trust is measured based on the user's frequent interaction patterns. Trust can be unidirectional or bi-directional depends on user behaviors. If a person has some prior information or recommendation about past experiences and performance, it increases the degree of trust in another person [161], [180]. As discussed in chapter 3, trust is classified as direct, indirect, and recommended. In direct trust (DT), a user directly trusts another person without any influence. In contrast, in the case of indirect trust (IDT), a user trusts another person indirectly by others' impact. In the case of recommendation trust (RT), trust is derived from other people's recommendations based on their experiences.

In this study, trust is considered a significant factor in establishing a relation between the users. Three collective psychological elements are considered Goodwill, Power, and Uprightness for measuring the user's fidelity. These three elements reflected the user's aggregate trustworthiness and replicated other users' willingness to establish a friendship. The pseudo-code for the trust score calculation is shown in algorithm 4.1.

Algorithm 4.1: Trust Score calculation Algorithm

Input: n number of users in the Social network.

Output: Trust Score (TS) of node i, $1 \leq i \leq V$

Steps:

1. $TS_i = \{0\}$;
2. for all node i ϵ {V} do

   Calculate Direct Trust (DT) of the node i

   $$DT(v_i) = \sum_{j \epsilon W_{(i)}} \alpha(n_j)$$

   Where $W_{(i)}$ is the set of neighbors of user i and α is the tunable balance parameters.

3. If ($DT(v_i) > k$) then

   $TS_1 = DT(v_i)$

   else

   $TS_1 = 0$

   end if;

4. Calculate In-Direct Trust (IDT) of the node i

   $IDT(v_i) = \sum_{j \epsilon W_{(i)}} [\alpha(n_j) + (1-\alpha) \sum_{k \epsilon W_{(j)}} C_k]$

   Where $W_{(j)}$ the set of neighbors of a neighbor of user i and C_k is the local clustering coefficient of node k.

4. if ($IDT(v_i) > l$) then

   $TS_2 = IDT(v_i)$

   else

   $TS_2 = 0$

   end if;

5. Calculate Recommendation Trust (RT) of the node i

   Suppose a user received m number of recommendations from other

   $RT(v_i) = \beta + \alpha \sum_{j \epsilon W_{(i)}} \frac{A_i}{\acute{M}_j^{out}} RT(v_j)$

   Here, β is the tunable balance parameters, A_i is the adjunct matrix of the graph G w.r.t. node i and $\acute{M}_j^{out}$ is the out-degree of node j.

6. if ($RT(v_i) > s$) then

   $TS_3 = RT(v_i)$

   else

   $TS_3 = 0$

   end if;

   end loop;

7. $TS_i = \sum_{i \in N}(TS_1 + TS_2 + TS_3)$

8. end**;**

In the above algorithm, three constant threshold variables k, l, and s are used such that (k, l, s) ϵ R and varies in the interval [0.5, 1) for all three types of trust, respectively. The threshold value of the entire three variables is network-specific and depends on network structure and dynamics. Once the trust is calculated, the next task is to calculate the Shapley value in a coalition. In a network, users made the coalition based on their shared interests, dyadic and triadic closure, strong and weak ties, and many other factors based on the network's nature [181], [182]. In this research, a new approach is proposed in which all possible combinations of the coalition of users are considered. Suppose the synergy of marginal contribution of the entire users in a coalition is higher than the amount of marginal contribution of users of any other group. All the users belong to that coalition considered as the opinion leaders in the network. In the same synergy, the coalition has the users with higher Eigenvector centrality considered for selection. The proposed approach is unique and has lots of potentials to present different variations based on various parameters.

In the real world, it happens that some of the players in the game have no much experience and cannot portray the game accurately. Most of the time, they are betrayed by their peers and colleagues. An inexperienced player may make decisions based on the cunning movement made by other players. In some situations, there is a probability that the other player may also be influenced by the other player's strategy up to some extent and modify their actions according to them. Thus, considering the above facts, four alternatives are considered for computing the individual payoff.

Solution 1: In this case both the player tries to persuade another player to accept or agree on their plan or change their strategy according to the plan of another player. Of course, each player is not readily getting the strategy of another player; therefore, there is a fixed payoff, called p, associated with each player, which is defined as follow:

p = +x if the other player changed their plan

p= -y if the other player unchanged their plan

p=-x if the player changed their plan

p= +y if the player unchanged their plan

Thus, the payoff table for both the player A and B is as follow:

Payoff for player A

	B unchanged plan	B changed plan
A unchanged plan	0	+y+x
A changed plan	-x-y	0

Payoff for player B

	B unchanged plan	B changed plan
A unchanged plan	0	-y-x
A changed plan	+x+y	0

Therefore, the payoff matrices for both the player A and B are as follow:

$$M_A = \begin{bmatrix} 0 & +x+y \\ -x-y & 0 \end{bmatrix} \qquad M_B = \begin{bmatrix} 0 & -x-y \\ +x+y & 0 \end{bmatrix}$$

In the above case, it is observed that if the player A unchanged their plan, player B have two option either change or keep their plan. In both cases, the payoff would be –x-y and 0, respectively. So this is the best option to choose 0 instead of –x–y, i.e., another player unchanged their plan. The above discussion concludes that both the player would not get any payoff and keep their plan unchanged and never reaching on an agreement. Thus, another solution is provided in which both the player agreed on an intermediate solution.

Solution 2: In this solution, an intermediate solution is suggested called 'agreement', i.e., both the players do not change their strategies and consent to an intermediate solution. In this case, the fixed payoff p has introduced for both the player as follow:

p= +i if the other player modify their plan towards the intermediate solution

p=-i if the player modify their plan towards the intermediate solution

Thus, the payoff table for both the player A and B is as follow:

Payoff for player A

	B unchanged plan	B changed plan	B agree
A unchanged plan	0	+y+x	+y+i
A changed plan	-x-y	0	-x+i
A agree	-y-i	+x-i	0

Payoff for player B

	B unchanged plan	B changed plan	B agree
A unchanged plan	0	-y-x	-y-i
A changed plan	+x+y	0	+x-i
A agree	+y+i	-x+i	0

Therefore, the payoff matrices for both the player A and B are as follow:

$$M_A = \begin{bmatrix} 0 & +x+y & +i+y \\ -x-y & 0 & +i-x \\ -i-y & -i+x & 0 \end{bmatrix} \quad M_B = \begin{bmatrix} 0 & -x-y & -i-y \\ +x+y & 0 & -i+x \\ +i+y & +i-x & 0 \end{bmatrix}$$

Again, in this case, It is observed that if the player unchanged their plan or agree on an intermediate solution, Player B has three alternatives, the first is to change the plan, the second is unchanged, and the third is to agree on an intermediate solution. In all the tree alternatives, the payoff for player B would be –y–x, 0, and –y–i, respectively. So the best option for player B is to choose payoff 0 and unchanged their strategy. It is analyzed that both the player's best choice is to unchanged their strategy and not receive any payoff. Thus, in this case, also both the player never ended with a proper solution and needed some better solution to obtain the optimal payoff. Therefore, another optimal solution is provided based on the centrality associated with the intermediate solution.

Solution 3: In this solution, a new parameter is introduced called 'distance', say d, which signifies the mean centrality between the users. In the social network, centrality plays a crucial role in defining the importance of a node. The centrality of the user depends on the network dynamics and network structure. As the dynamics of the network changes, the centrality of the user also varies over time. It is evident that if the centrality of a user is far above the ground, a user is reachable and accessible by most of the other users easily and may influence them with their behavior and technical knowledge. In the proposed approach, BC, CC, DC, and Clustering coefficient C_i to define the distance d_{ij} between the users i and user j in the network using the Eq. (4.3).

$$d_{ij} = \sum_{i \epsilon N-j}^{n} \sqrt[2]{\frac{BC_i * CC_i}{DC_i}} + \lambda C_i + \rho C_j \qquad \ldots (4.3)$$

In the above equation, both λ and ρ are the weighted coefficients which are used for balancing the distance d. the value of λ and ρ between [0, 0.5]. It is also found that, if the clustering coefficient of the user increases gradually, the distance between the users also increases. Now, the proposed payoff p for both the player A and B are as follow:

p = +x if the other player changed their plan

p = -y if the other player unchanged their plan

p = -x if the player changed their plan

p = +y if the player unchanged their plan

p = $+i+\frac{1}{d}$ if the other player modify their plan towards the intermediate solution

p = $-i+\frac{1}{d}$ if the player modify their plan towards the intermediate solution

Thus, the payoff table for both the player A and B is as follow:

Payoff for player A

	B unchanged plan	B changed plan	B agree
A unchanged plan	0	+y+x	$+y+i+\frac{1}{d}$
A changed plan	-x-y	0	$-x+i+\frac{1}{d}$
A agree	$-y-i+\frac{1}{d}$	$+x-i+\frac{1}{d}$	$\frac{2}{d}$

Payoff for player B

	B unchanged plan	B changed plan	B agree
A unchanged plan	0	-y-x	$-y-i+\frac{1}{d}$
A changed plan	+x+y	0	$+x-i+\frac{1}{d}$
A agree	$+y+i+\frac{1}{d}$	$-x+i+\frac{1}{d}$	$\frac{2}{d}$

Therefore, the payoff matrices for both the player A and B are as follow:

$$M_A = \begin{bmatrix} 0 & +x+y & +i+y+\frac{1}{d} \\ -x-y & 0 & +i-x+\frac{1}{d} \\ -i-y+\frac{1}{d} & -i+x+\frac{1}{d} & \frac{2}{d} \end{bmatrix}$$

$$M_B = \begin{bmatrix} 0 & -x-y & -i-y+\frac{1}{d} \\ +x+y & 0 & -i+x+\frac{1}{d} \\ +i+y+\frac{1}{d} & +i-x+\frac{1}{d} & \frac{2}{d} \end{bmatrix}$$

In this case, it is analyzed that if both the player agreed on the intermediate solution, the payoff for the player is $\frac{2}{d}$. Although this solution produced better results as compared to solution 3, again, the same problem occurred in which if any of the players unchanged their plan, another player may have three choices –y–x, 0, and –y–i+ $\frac{1}{d}$ to take the payoff. So the best option for another player is to choose payoff 0 if (y+i) > $\frac{2}{d}$ and agree on an intermediate solution. In the particular case, another player may receive the payoff of 0, if both the players changed or unchanged their plan simultaneously. So again, for overcoming this problem, a new solution is suggested along with the parameter d.

Solution 4: In the real-world, it has been observed that it is not necessary that each time a user completely disagrees with the other player strategy. There is a probability that up to what level a player convinces another player with their plan. In this solution, again, a new parameter initiated called 'u', which signifies the degree of inducement in the game. Two probability u_a and u_b are defined that signifies the probability that up to what level, player A influences player B, and player B, respectively, influence player A. So, in this game, the payoff of each player is controlled by two parameters u_a and u_b. If player A is not able to influence player B, then the (1- u_a) be the probability to receive the payoff. Now the payoff for both the player is defined as follow:

Payoff for player A

	B unchanged plan	B changed plan	B agree

A unchanged plan	$y((1-u_b)-(1-u_a))$	$y(1-u_b)+xu_a$	$y(1-u_b)+i+\frac{1}{d}$
A changed plan	$-xu_b-y(1-u_a)$	$x(u_a-u_b)$	$-xu_b+i+\frac{1}{d}$
A agree	$-y(1-u_a)-i+\frac{1}{d}$	$xu_a-i+\frac{1}{d}$	$\frac{2}{d}$

Payoff for player B

	B unchanged plan	B changed plan	B agree
A unchanged plan	$y((1-u_b)-(1-u_a))$	$-xu_b-y(1-u_a)$	$-y(1-u_a)-i+\frac{1}{d}$
A changed plan	$y(1-u_b)+xu_a$	$x(u_a-u_b)$	$xu_a-i+\frac{1}{d}$
A agree	$y(1-u_b)+i+\frac{1}{d}$	$-xu_b+i+\frac{1}{d}$	$\frac{2}{d}$

Hence, the payoff matrices for both the player A and B are as follow:

$$M_A = \begin{bmatrix} y((1-u_b)-(1-u_a)) & y(1-u_b)+xu_a & y(1-u_b)+i+\frac{1}{d} \\ -xu_b-y(1-u_a) & x(u_a-u_b) & -xu_b+i+\frac{1}{d} \\ -y(1-u_a)-i+\frac{1}{d} & xu_a-i+\frac{1}{d} & \frac{2}{d} \end{bmatrix}$$

$$M_B = \begin{bmatrix} y\big((1-u_b)-(1-u_a)\big) & -xu_b-y(1-u_a) & -y(1-u_a)-i+\frac{1}{d} \\ y(1-u_b)+xu_a & x(u_a-u_b) & xu_a-i+\frac{1}{d} \\ y(1-u_b)+i+\frac{1}{d} & -xu_b+i+\frac{1}{d} & \frac{2}{d} \end{bmatrix}$$

In this case, it is found that if both the players do not change their plan, still both the player may get some payoff based on the parameters u_a and u_b with the constraints that $1>u_b>0$, $1>u_a>0$. Similarly, if both the player agreed on the intermediate solution, the payoff may remain the same for all. Therefore the proposed solution motivates the player to plan their strategy and choose the next step accordingly. Suppose, at time t, both the player takes the decision based on the previous progress, the time progression needed to convince another player can be represented using Eq. (4.4)

$$E_A(t+1) = E_A(t) + \mu\,(E_B(t) - E_A(t)) \quad \text{if } 0 < \mu < 0.5 \qquad \ldots(4.4a)$$

$$E_B(t+1) = E_B(t) + \eta(E_B(t) - E_A(t)) \quad \text{if } 0 < \eta < 0.5 \qquad \dots(4.4b)$$

In the above equation, both μ and η both are the configurable variables that lie between [0, 0.5]. Finally, it is noticed that if $u_a < (i + \frac{1}{d})$, player A decides to take the decision' agreement' whatever the strategy is chosen by player B. Similarly, if $u_b < (i + \frac{1}{d})$, player B chooses the decision agreement, whatever the plan is decided by player A.

4.4 Calculation of synergy

A network's synergy is defined as the combination of nodes' ability or power that produces more value than the separately two or more nodes can do. A regular machine can do more productive work than another machine if all the machine components coordinate seamlessly with any mutual inference [178]. Likewise, in social networks, a coalition of a lesser number of organized people might produce more productive and innovative outcomes than a group of more disorganized people. The main phenomenon of synergy is based on structural holes in the network [183]. A structural hole is created when a single node connected with the vast network, and the same node also joined with another node related to an extensive network. These two nodes are the only medium through which information can pass from one group to another group. These nodes are responsible for sharing knowledge and communication [184]. Researchers suggested that these nodes have a high impact on network operations because they may simultaneously affect multiple links. In Fig. 4.3, consider the two networks having a node with a higher degree centrality (blue color node), and most of the other nodes connect with that node. If any node of a network needs to share information with another network, all the data might pass via this node, so this node plays a critical role in the synergetic coalition of networks.

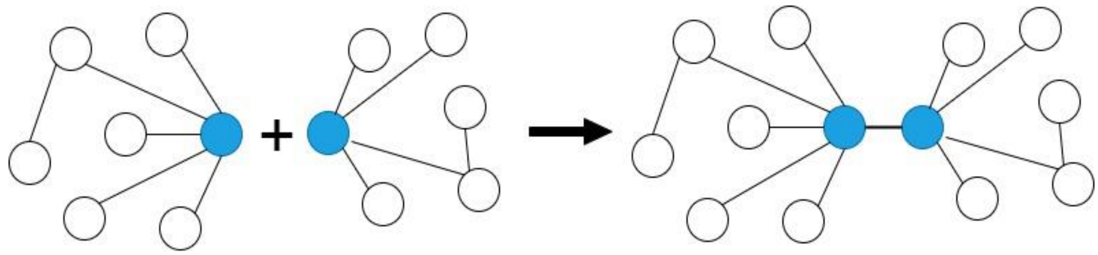


Fig. 4.3. A synergetic coalition of two networks

Two or more nodes are likely to be integrated when the merging outcome is hugely higher than expected. For measuring the synergy of the merger, we used Eigen centrality (EC) as the major component as the source node would attempt to merge with only those target nodes having a higher Shapley value. So the synergy ω_{ij} of the merger of two nodes and total synergy of a coalition can be measured using Eq. (4.5) and Eq. (4.6).

$$\omega_{ij} = \frac{1}{2}\{(EC)_i + (EC)_j\} * \frac{1}{2}\partial\{(SP)_i + (SP)_j\} \qquad \ldots (4.5)$$

$$\varphi = \sum_{i=1}^{x} \sum_{j=i+1}^{x} \delta * \left(\frac{\omega_{ij}}{x}\right)^{c} \qquad \ldots (4.6)$$

Where x is the total number of nodes in the coalition, c is the specific condition in which the user behaves, δ, and ∂ is the coordination rate that defines the coordination among the users. The value of c, δ and ∂ varies between 0 and 1. The flow chart of the proposed approach is shown in Fig. 4.4, and pseudo-code revealed in Algorithm 4.2.

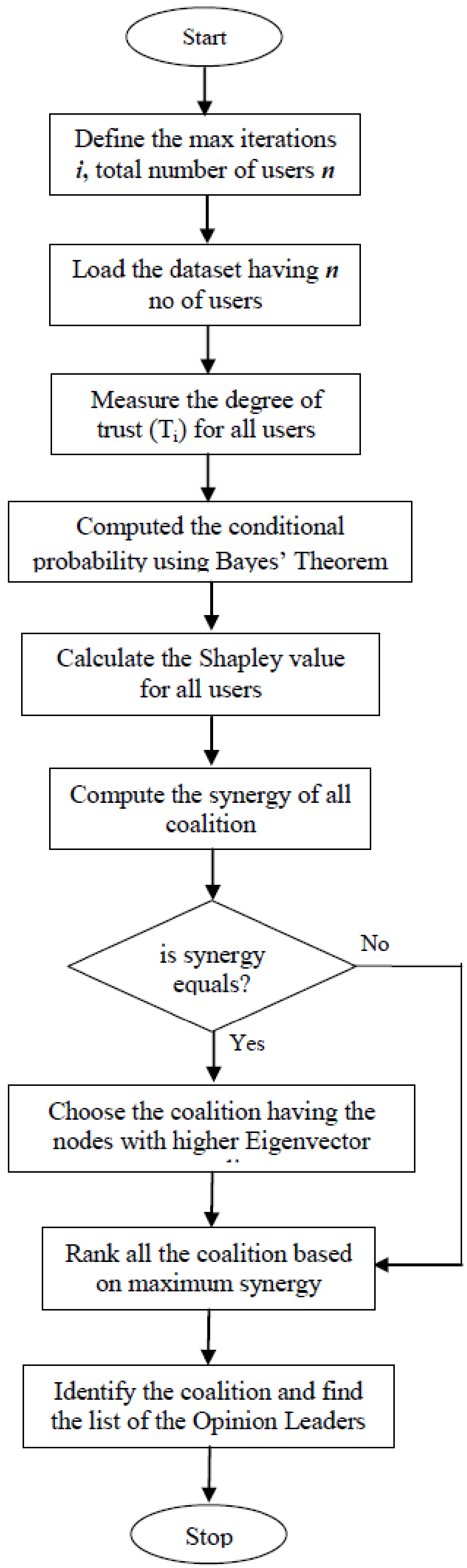


Fig. 4.4. Flow chart of the proposed Game theory-based algorithm

Algorithm 4.2: Game theory-based **O**pinion **L**eader **D**etection **(GOLD)**

Input: n number of users with their degree of trust

Output: a coalition of opinion leader

Steps:

1. Measure the degree of trust (T_i) for all nodes in the network

   S[] ←{0};

   for each node i ϵ {V} do

   for j=1 to V

   S[j] ←{T_i};

   end;

2. Computed the conditional probability (p_d) to connect with another node in trustor–trustee tree based on Bayes theorem

   $$p_d(\frac{T_i}{T_j}) = \frac{p_d\left(\frac{T_j}{T_i}\right) * p_d(T_i)}{p_d(T_j)}$$

   for j=1 to V

   P[j] ←{p_d};

   end;

3. Calculate the Shapley value of each user using Eq. (4.2).

   for j=1 to V

   SP[j] ←{SP_i};

   end;

4. Measure the user's marginal payoff using all the possible alternative solutions 1, 2, 3, and 4.
5. Compute the synergy (SE) of all coalition using Eq. (4.5).
6. If the synergy of two coalitions is equal, choose the coalition having the nodes with higher Eigenvector centrality.
7. Rank all the coalition based on maximum synergy.
8. Identify the top coalitions and find a list of opinion leaders.

4.5 Complexity of the algorithm

The algorithm's complexity determines the total amount of time and memory needed for the implementation of the entire program for specified input. Here, two unique algorithms are anticipated; the former is addressed to calculate the degree of trust, and the later depicted the coalition's identification in the network. In the first algorithm, the direct, indirect, and recommended trust is measured based on goodwill's fundamental behavioral attributes. For measuring the direct trust, all nodes are considered directly connected with the root node. Hence the complexity of the direct trust is O(E). Next, for measuring the indirect trust, the depth-first search is used to measure the clustering coefficient of the node. Thus, the time complexity of indirect trust measurement is O(V+E). In the case of a recommended trust, the node's adjacency matrices are used to measure the recommendations suggested by its neighbors. So, the complexity of the recommended trust is O(V+E). Therefore, the overall time complexity of the algorithm is O(V+E)+O(E)≈O(V+E).

In the case of game theory-based opinion leader detection, initially, the conditional probability of the nodes is calculated in the network based on Bayes' theorem. In this case, all the users are independent of each other; the action or decision taken by them is to depend on their acquaintance's actions. Therefore the time complexity of this procedure is O(V). Next, the Shapley value of each user is calculated in the coalition. The complexity of Shapley value calculation depends upon the type of input, i.e., how the input is given in the problem domain. As discussed earlier, it is assumed that either the user belonging to the same coalition or is contiguous to as a minimum x-neighbors who are in the coalition. Therefore, the time complexity to evaluate the Shapley value of each node is O(|V|+|E|).

Further, a calculation is performed on the whole possible number of permuted coalitions to measure the probability that which permutation produces the highest synergy. The Shapley value and Eigenvector centrality are used to measure the synergy of each coalition. Thus, the time complexity to measure the synergy for each coalition is $O(|V|^2)$. Next, all the coalitions are ranked based on their synergy. This step's complexity is based on the total number of coalitions, say e, identified in the network. So the time complexity of this step is O(e). Thus, the overall time complexity of the proposed method is $O(|V|+|E|) + O(|V|^2) + O(e) \approx O(|V|^2)$

4.6 Experimental analysis and results

The two real networks are used; Wiki-vote and Bitcoin OTC trust weighted signed dataset. A network visualization tool, Gephi 0.9.2, is used for measuring the network parameters and analysis purposes [185]. Python 3.0 and Intel i7 multi-generation processors are used to obtain the experimental results [186]. Now, the detailed description of the datasets is as follows.

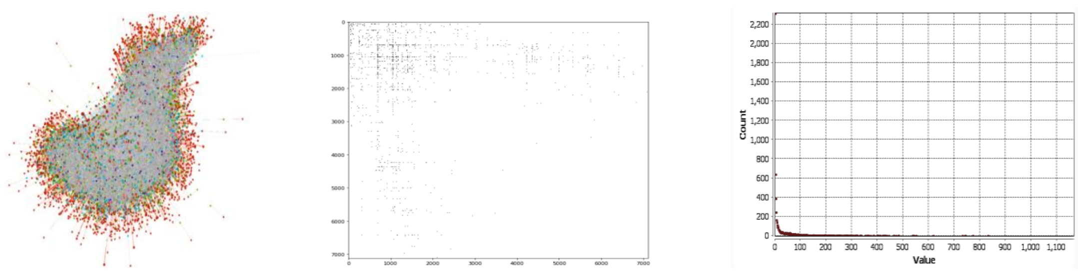
4.6.1 Datasets

4.6.1.1 Wiki-vote social network

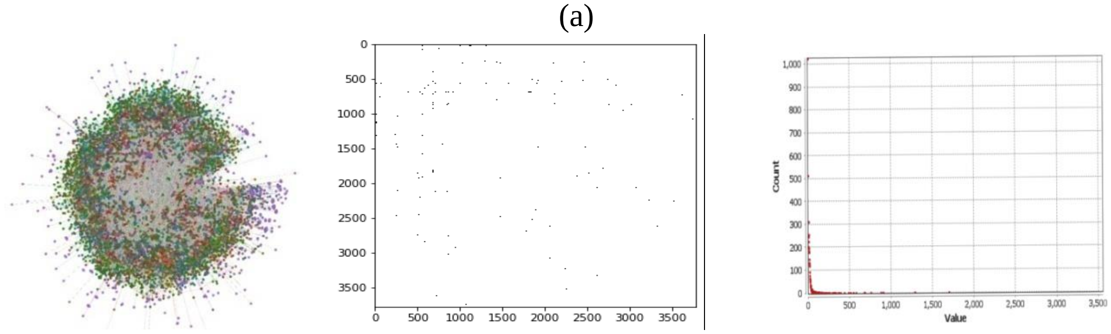
Wikipedia is a free information portal that provides various types of information for its users. Wikipedia allows the facility for its user to add new information and features under the supervision of administration. If the user wants to update any technical details on an issue, a request is delivered to the administrators, who have the right to update the information. If a user wants to become an administrator, an application of adminship gave to Wikipedia authority. The election of an administrator is either done by the community users or by the voting mechanism. The dataset contains all the voting information, including the total number of votes and the number of candidates participating in the election. The node in the dataset indicates the total number of users, while an edge from one user to another user depicts the user's voting choice. The dataset has a total of 7115 nodes and 103689 edges. The density and average clustering coefficient of the network is 0.002 and 0.081, respectively. Some structure holes are valuable to make a bridge between two or more huge clusters [187].

4.6.1.2 Bitcoin OTC trust weighted signed network

Bitcoin OTC trust weighted signed dataset is a trust-dependent network in which users trade the Bitcoin using a platform called Bitcoin OTC (Over-the-Counter) [188]. During Bitcoin trading, the customer's identity is hidden, so it is required to maintain a database of user's reputations to avoid any counterfeit and hazard. Such a network, also known as a who-trust-whom network, in which the user grades another user on the scale of -10(total distrust) to +10(absolute trust). It formed the web-of-trust network that has a total of 5881 nodes that represent the users and a total of 35593 links that describe the user's trust grading on another user. The density and average clustering coefficient of the network is 0.002 and 0.267, respectively. The geographical structure, degree centrality, and adjacency matrix of both the network are shown in Fig. 4.5.



(a)



(b)

Fig. 4.5. Geographical structure, adjacency matrix, and degree distribution of (a) Wiki-vote network, (b) Bitcoin OTC trust weighted network

4.6.2 Analysis and visualization of the experimental result

Now, the proposed algorithm is deployed on both the dataset. In the real world, it is found that trust is not considered as a discrete value, i.e., not in binary form (yes or no). So in the first step, the trust is measured among the user based on Algorithm 4.1. Initially, some degree of trust is assigned to each user based on their experience and knowledge. Further, a trustor-trustee tree is designed by randomly select a seed user and its subsequent neighbors iteratively. For measuring the trust, some parameters are used, say α and β, to obtain the optimal output. The value of parameters is decided based on experimental analysis.

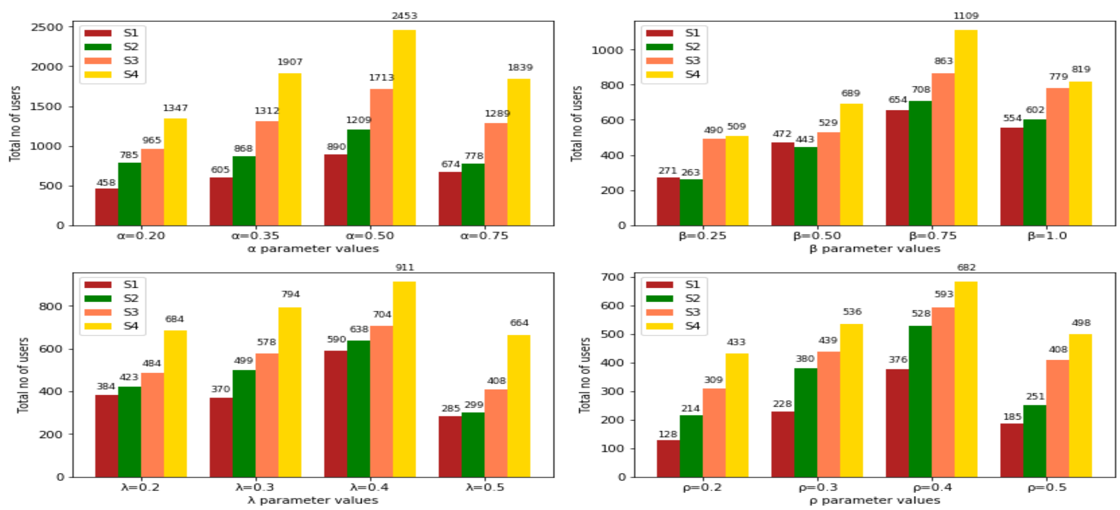


Fig. 4.6. Parameters value estimation for Wiki-vote dataset

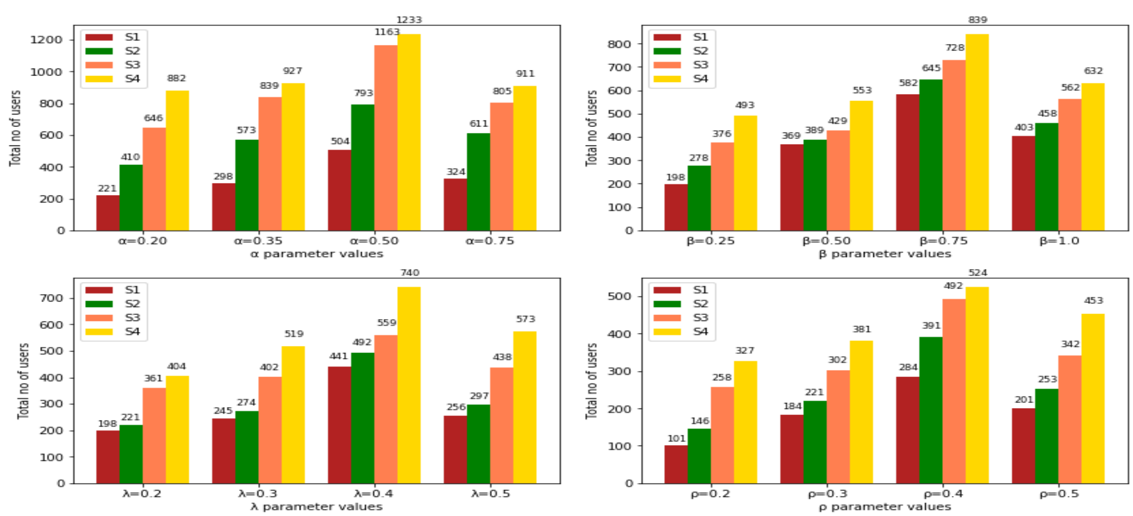


Fig. 4.7. Parameters value estimation for Bitcoin OTC trust weighted dataset

In Fig. 4.6 and Fig. 4.7, we found that at α=0.5, β=0.75, λ=0.4, and ρ=0.4, the proposed algorithm-generated only those opinion leaders who are significantly influencing the followers. Similarly, two weighted coefficients λ and ρ, whose value is also identified based on experimental analysis, measure the distance between the users.
Next, the conditional probability between the users is measured using Bayes' theorem. To implement the game theory approach in the network, 100 units are initially assigned to each user. Four alternative solutions have projected to calculate the marginal contribution of each player in a game. Still, the fourth solution produced better results and took a lesser number of iterations as compared to other solutions. In Table 4.1, the top-5 coalitions are obtained from both the datasets and the total number of opinion leaders with their average marginal payoff.

Table 4.1. Top-5 coalitions in the Wiki-vote and Bitcoin OTC trust weighted dataset

Data set	Coalition	Total number of opinion leaders	Collective Synergy
Wiki-Vote Dataset	C1	382	278.45932
	C2	478	275.06345
	C3	684	270.48298
	C4	952	264.95834
	C5	1689	262.10047
Bitcoin OTC trust weighted dataset	C1	205	186.99605
	C2	380	185.26755
	C3	538	183.33674
	C4	884	180.38609
	C5	1087	176.19118

In the social network, there are also some other SNA centrality measures to find the prominence of the node. In Tables 4.2 and 4.3, we illustrated the Shapley value of top-10 opinion leaders of the coalition having maximum marginal payoff and compare their ranking with other centrality measures.

Table 4.2. Top-10 opinion leaders in Wiki-Vote data set based on Shapley value

Rank	Node id	Shapley Value	DC	CC	BC	EC	PR
1	3542	97.53765	0.24915	0.48009	0.32025	0.27582	0.41251
2	267	97.52764	0.24832	0.46276	**0.32586**	0.27362	**0.41589**
3	1089	97.52116	0.24786	**0.48769**	0.31254	0.26845	**0.42185**
4	5376	97.51986	0.24641	**0.46388**	**0.32817**	0.26521	0.40356
5	1008	97.51912	0.24465	**0.47894**	**0.32118**	**0.26559**	**0.41664**
6	387	97.51854	**0.24562**	**0.47106**	0.31958	0.26331	**0.41985**
7	6959	97.51606	0.24378	0.46228	**0.32682**	0.25487	**0.42581**
8	2968	97.51594	0.24371	**0.46834**	0.31584	**0.26580**	**0.41607**
9	4822	97.51568	**0.24374**	0.45732	**0.31958**	0.25148	**0.40931**
10	3911	97.51561	0.24366	**0.46117**	**0.31705**	**0.26958**	**0.41583**

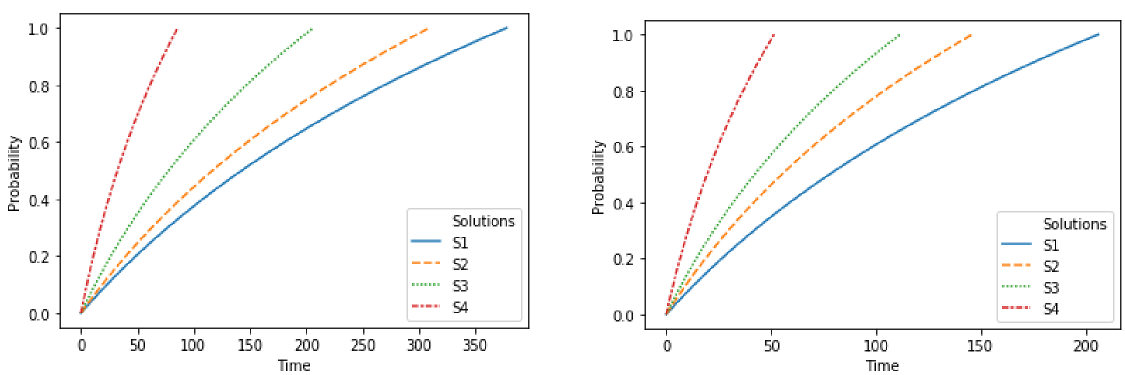
Table 4.3. Top-10 opinion leaders in Bitcoin OTC trust weighted dataset based on Shapley value

Rank	Node id	Shapley Value	DC	CC	BC	EC	PR
1	2829	93.34872	0.32514	0.29201	0.36128	0.22183	0.41006
2	74	93.34773	0.31522	**0.29651**	**0.36982**	0.22058	0.41002
3	1008	93.34754	**0.33201**	**0.29536**	0.36108	0.22015	0.40995
4	3116	93.33543	**0.32584**	0.28105	**0.36471**	**0.23197**	**0.41264**

5	1427	93.31008	0.31998	0.27847	0.35927	**0.23496**	0.40865
6	4275	93.30635	0.30584	**0.28853**	0.35461	0.22004	**0.41106**
7	2731	93.30223	**0.32115**	**0.29147**	**0.36014**	0.21986	0.40137
8	694	93.30018	**0.31458**	0.27651	0.35100	0.21547	**0.41042**
9	3952	92.29474	**0.31025**	0.27502	0.34925	**0.22010**	**0.40892**
10	4937	92.29223	**0.32264**	**0.27984**	**0.34998**	0.21327	**0.41072**

The above tables show that the highest Shapley value user does not have the highest degree centralities. Only a few users hardly maintain the centrality hierarchy, and the rest of the users are scattered in the record as per the ranking of the Shapley value. Therefore, it does not mean that the node with the higher centralities also has a higher Shapley value. It depends on the trust, payoff, probability of coalition, and network structure.

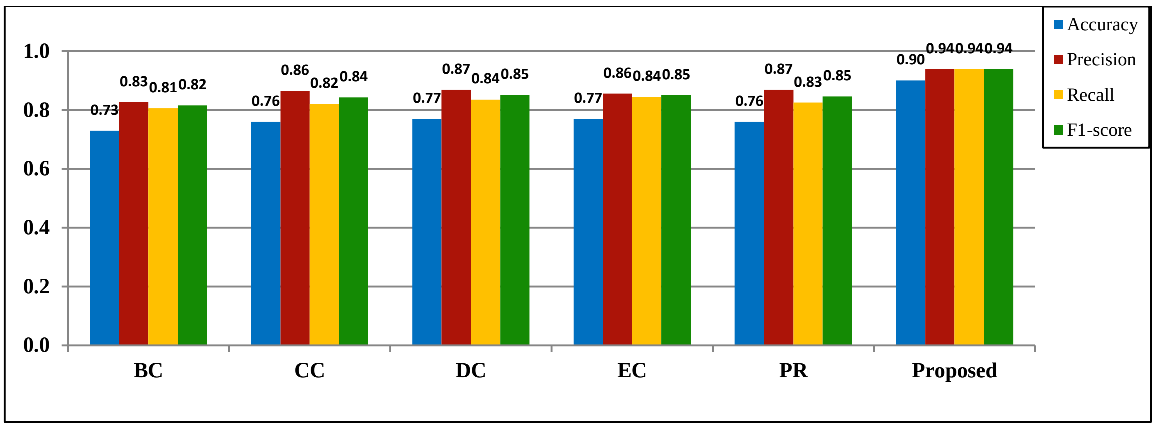
During the evaluation, it is identified that solution 4 converges earlier as compared to other alternative solutions, as shown in Fig. 4.8. So, it can be concluded that solution 4 produced the desired outcome in a smaller number of iterations and proven its significance to implement the proposed algorithm. Thus, the entire obtained coalition and corresponding opinion leaders evaluated using solution 4.



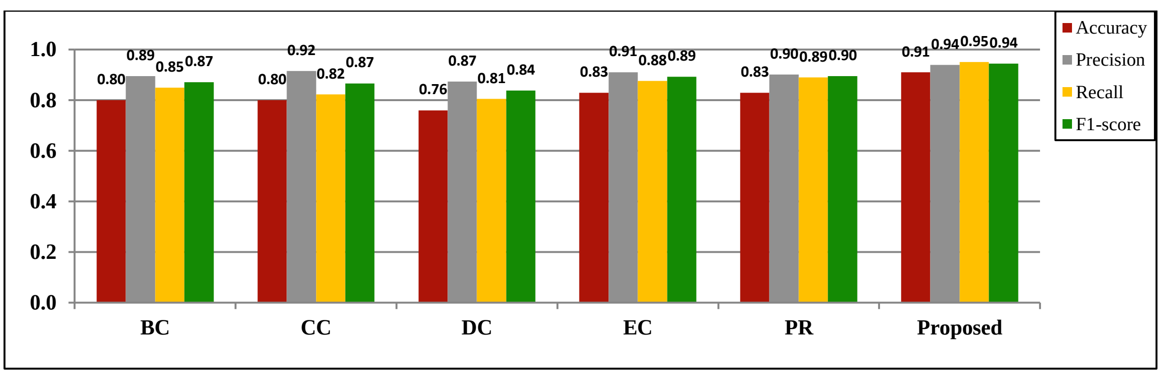
(a) (b)

Fig. 4.8. Rate of convergence for all the proposed alternatives to calculate payoff for (a) Wiki-vote network, (b) Bitcoin OTC trust weighted network

Further, to justify the superiority of the proposed algorithm, the experimental results are compared with the standard SNA measures based on accuracy, precision, recall, and F1-score [189]. The outcomes are also compared based on execution time and power of influence. Most of the researchers compared their findings with the standard SNA measures, and only a few compared the outcomes with the other methods based on some common features in a specific domain. There is a need for True Positive, True Negative, False Positive, and False Negative [81] observations for evaluating these measures. The compared results are based on the mentioned performance metrics shown in Fig. 4.9.



(a)



(b)

Fig. 4.9. Comparative analysis to the accuracy, precision, recall, and F1-score with other SNA measures for (a) Wiki-vote dataset (b) Bitcoin OTC trust weighted dataset using the game-theory approach

The above analysis can quickly examine that the proposed methodology conferred better results over other SNA measures for both datasets. The proposed approach obtained an approximate 90% accuracy and having 94% precision for the wiki-vote dataset. Similarly, for the Bitcoin OTC trust weighted dataset, the accuracy is around 91%, and precision is nearly 94%. Besides, the total implementation time is also examined needed to deploy the entire algorithms on the dataset. It is found that the suggested approach needed lesser time and about 37% reduced in contrast to other SNA measures, as shown in Fig. 4.10.

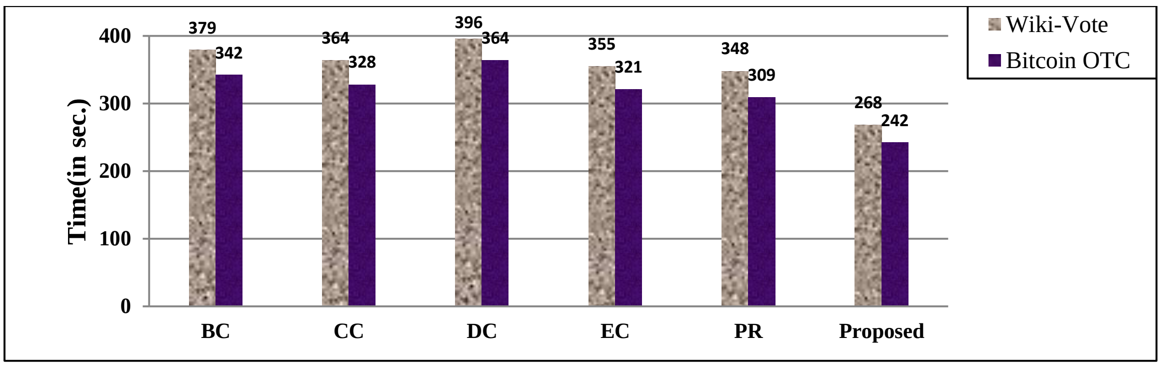


Fig. 4.10. Comparison of execution time for Wiki-vote dataset and Bitcoin OTC trust weighted dataset with other SNA measures using game theory approach

One of the opinion leaders' significant contributions is to diffuse the products over a while [190]. The degree of diffusion depends on the total number of opinion leaders involved in the process, the total number of followers, and the time at which the revolution took place. The main task in diffusion consists of accepting any modification or innovation in the product by the customers in the real world [92]. Sometimes the company has to struggle many of the years to promote its products in the commercial market, still not able to convince the consumers. So, an opinion leader is too supportive of disseminating products and resolving the cold start problem in the recommended system. Moreover, it is observed that the proposed approach required etiquette number of opinion leaders for both the dataset. Simultaneously, other SNA measures need comparatively more opinion leaders to achieve full adoption by the customers, as shown in Fig. 4.11.

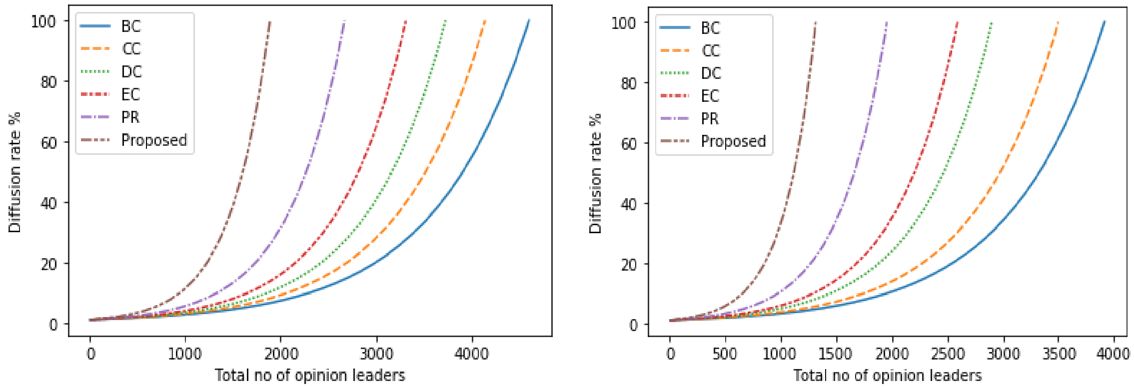


Fig. 4.11. Impact of the total number of opinion leaders on the diffusion rate in (a) Wiki-vote dataset (b) Bitcoin OTC trust weighted dataset

It is analyzed that the proposed method needed a total of 1859 opinion leaders for the Wiki-vote dataset and 1304 opinion leaders for the Bitcoin OTC trust weighted dataset while the other SNA measure needed additional opinion leaders for the diffusion of products and services entirely in the social network.

4.7 Chapter summary

In this chapter, a game theory-based approach is explained to find the coalition of opinion leaders based on the group's maximum collective synergy. Shapley value is also exploiting to find the average marginal contribution of a user in a coalition utilized to produce synergy. A user probably interacts with the others based on the degree of trust, and likelihood conditional probability depends on the other user's action. The payoff of each user is also sensibly calculated in every coalition by providing the four alternative solutions. Therefore, the social networks' power is uniquely combined with the logic of game theory and trust and conditional probability to precisely identify the opinion leaders in the social network. The proposed approach is also efficient compared to other SNA measures and produced the enhanced results with around 90% accuracy and 94% precision. The computation time of the algorithm is also reduced by 27% as compared to other SNA measures.

The suggested approach works as an intelligence expert system that intelligently identifies opinion leaders' lists using the game theory approach in the social network. Nowadays, most industries adopt opinion leaders to promote their commodities inside the real world and use their expertise to analyze the reviewers' and other promoters' feedback. So, they are very supportive of increasing the wholesale price and gross growth rate of the product. In a nutshell, an opinion leader's role is genuinely deserving and prominent for developing new products and can persuade the assessment by their convincing power and strategy. In the future, the power of other centrality measures and game theory approaches may be combined to find the list of opinion leaders. Besides, the amalgamation of computational intelligence-based techniques, social network dynamics, and the evolutionary game theory approach may detect the promising opinion leader in social networks [191], [192].

Chapter 5

Opinion Leaders for information diffusion using Graph Neural Network

The Graph Neural Network (GNN) is a deep learning-based model that modernized neural networks' efficiency by analyzing and extracting the latent dependencies and confined embedding via message passing and neighborhood aggregation of data in the network. In this chapter, an exclusive GNN for Opinion Leader Identification (GOLI) model proposed utilizing the power of GNN to categorize the opinion leaders and their impact on the online social network. Section 5.1 explains the overview of the chapter. The introduction of GNN is discussed in section 5.2. Section 5.3 elaborates on the proposed GOLI model in detail. In this model, the n-node neighbor's reputation of the node is measured based on materialized trust. Centrality conciliation is performed instead of the conventional node embedding mechanism. Section 5.4 discussed the experimental results performed on six different online social networks consisting of billions of users to validate the model's authenticity. Section 5.5 demonstrated how the opinion leaders effectively used information diffusion through various performance metrices. Finally, section 5.6 concluded with the chapter abstract.

5.1 Overview

The deep learning-based model achieved more sensation and strengthened the research in image processing and restoration, natural language processing, computer vision, speech recognition, healthcare, finance, automobiles, bioinformatics, defense, grid computing, and many more [193]–[195]. Various machine learning tasks such as object identification, classification, regression, outlier detection, and many more have modernized by different stem-to-stern deep learning archetypes such as Artificial neural network (ANN), Recurrent

Neural Network, Convolution Neural Network(CNN), Autoencoders, Long-Short-Term-Memory(LSTM) and many more [196]. The advanced electronic circuit, i.e., GPU (Graphical processing unit), speedily and updates memory using parallelization to step up the generation of image, text, and video in a frame buffer, also one of the prominent causes behind the exposure of deep learning models. Besides, the ability to handle and process a large amount of training data, capture the multi-scale confine latent features, and map them into Euclidian space strengthens the deep learning-based model's power. In general, the deep learning-based model having six main components: raw input, number of multiple hidden layers, shared weights, activation functions (Sigmoid, ReLU, etc.), loss function (Gradient Decedent, etc.), and output (Binary Classification, Multi-class Classification, Regression, Clustering, etc.) [197], [198].

5.2 Graph neural network

The GNN is ubiquitous and pervasive to depict deep neural networks on graph structure data. Naturally, the developed deep learning-based models are not so much capable of managing and optimized graph data. For example, CNN is well building for structured grid data, while the RNN sounds suitable for time series data and progression. GNN based model contains the integration of both of these frameworks without any geographical constraints. Usually, in the previous build-up approached, the adjacency matrix or variations used as input data. The adjacency matrix's primary issue is the dependency of the arrangement of nodes. If we consider the isomorphic graph and evaluate it, the different forms of adjacency matrix give the same network's other output. The first primary problem with the graph data is that it can not be represented in Euclidian space, i.e., it can not be depicted in any coordinate system [199].

The second problem is with the graph is that they do not have a fixed organization because of isomorphism. One of the foremost benefits of GNN is to produce the same results for all possible permutations. The leading function of the GNN is to identify the node embedding of each node by considering the information received from its neighbors. In

Fig. 5.1, it is observed that different nodes can be embedded in n-dimensional Euclidian space through respective feature vectors. A euclidian space or vector space is n-dimensional (usually 2-dimensional or 3-dimensional) space to store x-points of real numbers. Various graph embedding techniques such as Locally Linear Embedding (LLE), HOPE, Word2vec, DeepWalk, Node2vec, Structural Deep Network Embedding, Autoencoder, and many more used for capturing node's state information [200]–[203] The state of the node is updated according to the neighbor's participation. A local transition function is distributed mutually among all the nodes, accountable for the state modification, while the local output function defined the produced outcomes.

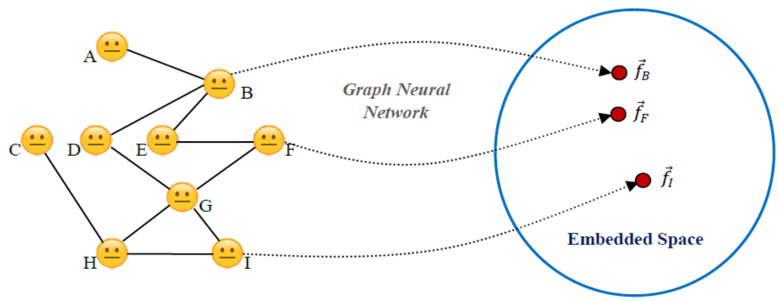


Fig. 5.1. Graph implantation in embedding space using GNN

During the message passing phase, the node's network embedding message transformed and updated the node's state according to the node's neighborhood, as shown in Fig. 5.2. All the corresponding neighbors aggregated and updated the embedding information as received from a node.

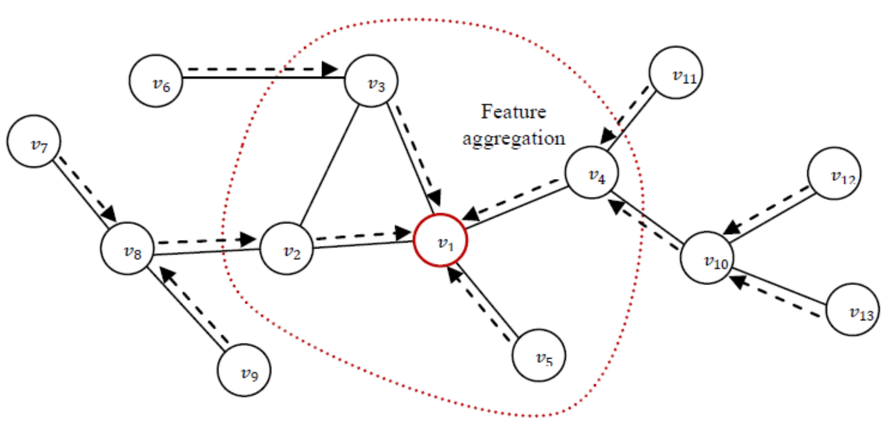


Fig. 5.2. Feature aggregation by a node in GNN

Therefore, technically, in the conventional GNN, the feature vector of a node x at (l-1)-th layer is represented by $\vec{f}_x^{l-1}$, modified using the cumulative feature vectors collected from the neighborhood set $И(x)$ of node x, and probably the weight $w_{x,y}^{l}$ received from neighbor y. The aggregation function is repeated up to the L layer, and the feature vector of (l-1)-th layer becomes the input of the l-th layer. Hence, after the total L(1,2,3,…,l) number of repetitions, the node x received the L-th order feature vector representation via their neighbors as shown in Fig. 5.3. The feature learning representation of node x at layer l is indicated as following Eq.(5.1) and Eq.(5.2).

$$\vec{f}_{И(x)}^{l} \leftarrow \text{Transformation}^{l}\left(\text{Aggregation}\left(\left[\vec{f}_x^{l-1}, w_{x,y}^{l}\right] \| \, y \in И(x)\right]\right)\right) \qquad \ldots (5.1)$$

$$\vec{f}_x^{l} \leftarrow \text{Join}\left(\vec{f}_{И(x)}^{l} \cup \vec{f}_x^{l-1}\right) \qquad \ldots (5.2)$$

The aggregation function can be min-max or average pooling function. The transformation is a model-driven operation that performed some non-linear transformation-based input received from (l-1)-th layer and few non-linear variables, i.e., σ(.). Join is a union operation used to combine the feature vector of a node x and collective neighborhood depiction.

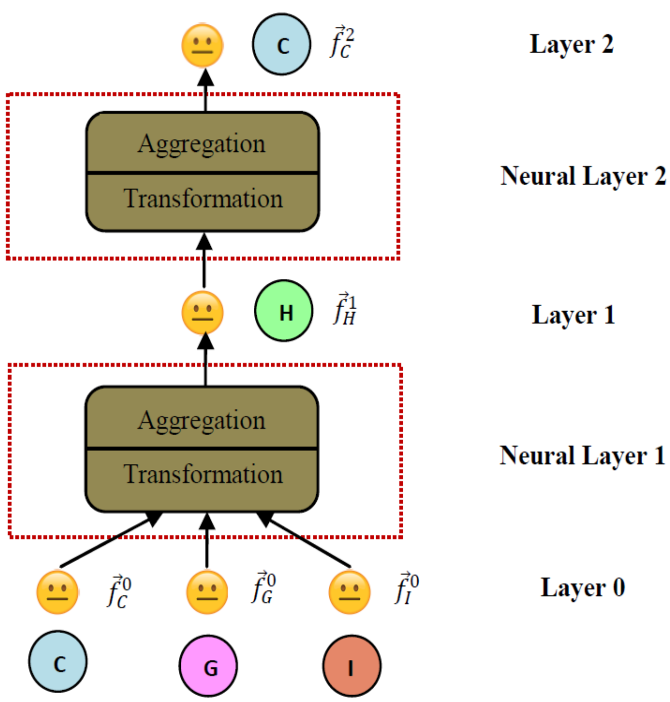


Fig. 5.3. The layered architecture of GNN

5.3 Proposed GOLI model

In this segment, the proposed GNN for Opinion Leader Identification (GOLI) model is discussed online. Initially, the motivation and key factors behind the generation of the model have been discussed. Next, the proposed model architecture explained that it consists of trust calculation, centrality reconciliation, and reputation aggregation. The training mechanism of the model is also demonstrated in detail.

5.3.1 Motivation

It is analyzed and found that the following aspects stimulate the opinion leader identification process.

- **Centrality knowledge:** A node with high centrality is considered the most powerful node in the network with no other measure. However, this is not the right way for assortment, yet one of the essential factors for the opinion leader selection. Centrality defines the power, connectivity, and influence of a node in the network.
- **Neighborhood consciousness:** A node is connected with multiple nodes in the network directly or indirectly. The neighbors of a node are directly communicated with the node and share state features. Dyadic and triadic relations are formed due to the neighborhood's homophily properties, i.e., the node's affinity to make a connection with the nodes having analogous attributes.
- **Mutual trustworthiness:** Trust is one of the well-known factors for the construction and development of the network. Connections between the nodes break or become weak due to the lack of trust. So, it is vital to consider and measure the degree of trust among the nodes. A trust can be direct, indirect, or recommended, depends upon the mutual connection and experience.
- **Utilization of reputation:** Besides network structure, node reputation is also needed and provided lots of information to find opinion leaders. The reputation of the node signifies the prominence, power, and strength of the node.

- **Agile adjustment:** In the real world, different kind of social network exists because of their working style, structure, content delivery, and users. So, it is essential to make such a type of model that could be easily adaptable with most of the social network and may utilize their characteristics effectively.

Therefore, the proposed model's primary target is to accomplish all the mentioned requirements as much as possible. The general building block of the proposed GOLI model is shown in Fig. 5.4. Initially, the feature vector matrix of the nodes is estimated. The feature vector matrix comprises the latent features derived from the network topology and node neighbors' information. The feature vector's value varies from network to network because of different kinds of working methods and topology. Next, the anticipated trust network is formed based on the feature vectors and trust. Here, trust plays a critical function in developing a network. The leading cause behind the generation of trust networks is to predict the degree of confidence and trustworthiness unceremoniously among the nodes to sustain the network's decision-making process. Next, the proposed technique is applied to calculate the reputation of the node. Finally, the top-n opinion leaders are elected based on the uppermost reputation score.

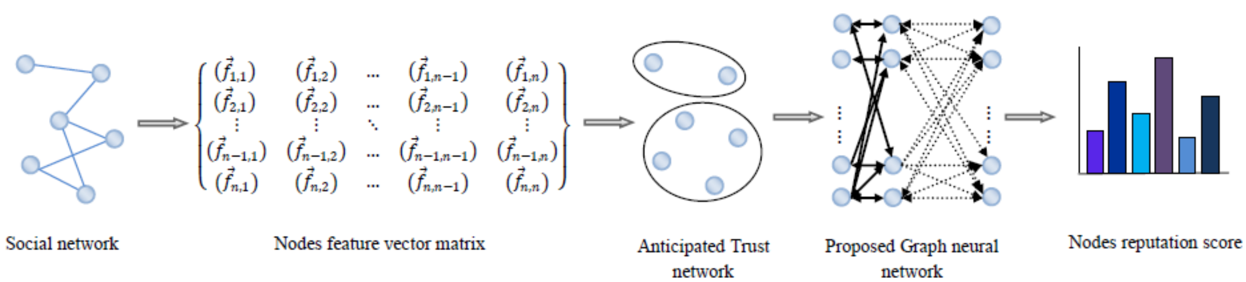


Fig. 5.4. The building block of the proposed GOLI model

In this work, few symbols are used for experimental analysis and model generation. The used characters, along with their description, are shown in Table 5.1.

Table 5.1. List of symbols along with their description used in GOLI model

$\mathcal{R}$	Set of nodes with a recognized reputation score

$\overrightarrow{f_x}$	The feature vector of node x
L	Total number of reputation aggregation layers
И(x)	Total numbers of first-order neighbors of node x
r_x^η	Reputation score of node x by Reputation Aggregator (RA) in η -th layer
L_η	η-th level of the reputation aggregation layer
r_x^*	Centrality reconciliation score of node x
d(x)	In-degree of node x
d' (x)	**P**reliminary centrality of node x
$T_{x,y}^\eta$	Node x trust on node y calculated by RA at η-th layer
$\vec{\delta}_\eta$	Attainable parameters used by RA at $\eta - $ th layer to measure $T_{x,y}^\eta$
$\beta_\eta,\ \gamma_\eta$	Learnable parameters used for tuning and normalization at η-th layer

5.3.2 Structural design of the model

The basic GOLI model architecture consists of a solo reputation aggregation layer(i.e. l-th layer) trailed by the centrality reconciliation module, as shown in Fig. 5.5. Initially, the node's feature vector is computed based on the various attributes depending on the network's nature and services. The feature vectors ensembling the latent features which grant more embedding knowledge of a node. Next, the feature vectors' information is utilized to measure the initial reputation of the node by its neighbors at layer zero. As soon as all the nodes received the information about their neighbors' initial reputation, they anticipated a trust network. Although the anticipated trust network's construction is merely based on user experience, social persuade privacy, and social interaction with other nodes. Next, the single reputation aggregation layer cumulated the reputation score and the measured trust received from multiple numbers of neighbors. The same instructions are followed by all the layers simultaneously for all the nodes. Hence the entire procedure supports parallel computation.

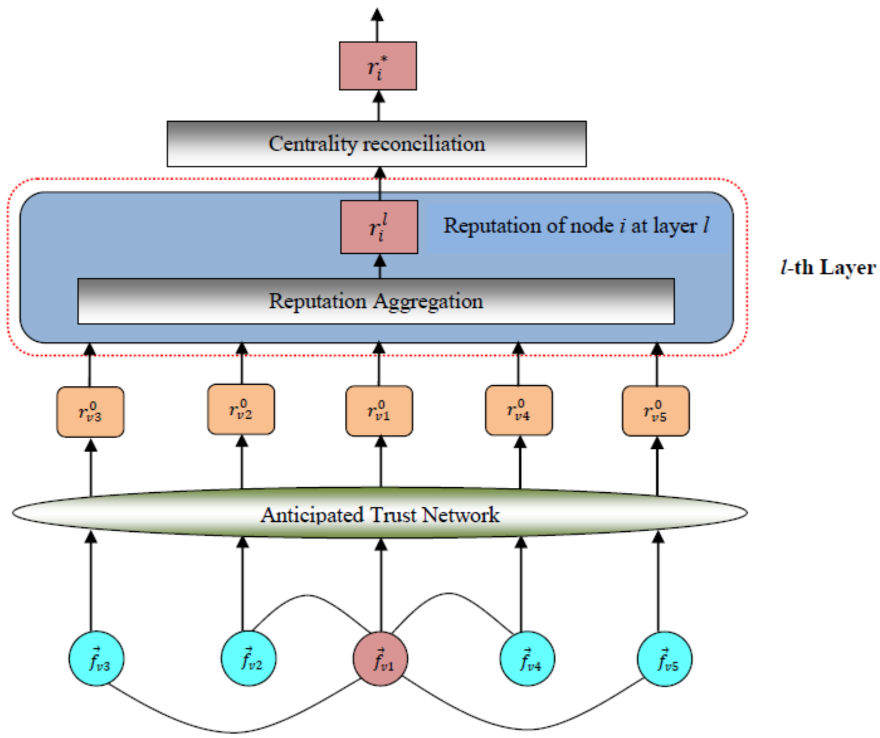


Fig. 5.5. The layered architecture of the GOLI model

Now the whole procedure is explained along with operational notations empirically. Let L be the total number of aggregation layers and L_η be the η-th number of the reputation aggregation layer. Let, $r_x^{\eta-1}$ The reputation score of node x by reputation aggregator in (η-1)-th layer that is to be transmitted to cumulate the reputation score at the η-th layer. At the initial layer, the reputation score r_x^0 of the node measured by applying the Artificial Neural network (ANN) on the input feature vector $\vec{f}_C^0$. The reputation score produced by the preceding layers (L>0) is used for calculating the mean of reputation score in the succeeding layers. Therefore, each layer's reputation score's importance also has a strong influence on the aggregate reputation score estimation. The reputation aggregator also considered the trust $T_{x,y}^\eta$, obtained from its corresponding neighbors for the computation of reputation score.

5.3.2.1 Centrality reconciliation

Centrality is one of the major components that affect the attractiveness of the node in the network. A node with higher centrality captures the center of attraction and impacts other nodes' influencing power. Generally, multiple centralities measures, such as degree, betweenness, closeness, eigenvector, PageRank, etc., has high gratitude for critical leaders recognition [87]. So, the preliminary centrality d'(x) of node x is described in Eq.(5.3).

$$d'(x) = \log(d(x) + \varepsilon) \qquad \dots (5.3)$$

Here, ε is a constant used for scaling and normalization. Further, the node centrality has an active impact on the opinion leader identification, so the extended and updated in-degree centrality $d^*_\eta(x)$ is used for the final reputation score calculation. The extended centrality of node x measured by score aggregator r^η_x at η-th layer is represented in Eq.(5.4).

$$d^*_\eta(x) = \beta_\eta * d'(x) + \gamma_\eta \qquad \dots (5.4)$$

Here, β_η and γ_η both are the learnable parameters used for tuning and optimization purposes. This centrality parameter is also utilized for the final calculation of node reputation score.

5.3.2.2 Trust measurement

The proposed GOLI model mainly focused on neighbor trustworthiness. Trust also plays a lead role in opinion leader identification in the social network [141], [204]. Even in the real world, only a trusted person can gain the other's person favor and support to disseminate a particular thing. Trust mainly depends on direct and indirect interactions. Sometimes, recommended trust also significant for the decision-making process. GOLI model follows the one-order trust calculation and utilizes the initial reputation value that a node gained from direct neighbors and their neighbors. At each layer, the trust between the nodes is calculated by the Reputation Aggregator (RA). So, the trust of node x on node y, i.e., $T^\eta_{x,y}$, calculated by RA at η–th layer represented in Eq.(5.5).

$$T_{x,y}^{\eta} = \frac{\exp\left(\alpha_{\delta}\left(\sum \vec{\delta}_{\eta}^{\,T}\left(r_x^l \,|\times|\, r_y^l\right)\right)\right)}{\sum_{z \epsilon И(x) \cup \{x\}} \exp\left(\alpha_{\delta}\left(\sum \vec{\delta}_{\eta}^{\,T}\left(r_x^l \,|\times|\, r_y^l\right)\right)\right)} \qquad \dots (5.5)$$

Where, $\vec{\delta}_{\eta}$ is a power vector used by RA at η-th layer while α_{δ} is non-linearity. r_x^l, r_y^l, and r_z^l depicts reputation score of node x, node y, and node z respectively by reputation aggregator in η -th layer. $|\times|$ operator indicates the multiplication operation between two reputation aggregation scores.

5.3.2.3 Reputation aggregation

The reputation aggregation score is used to find the importance of a node in the network. The proposed model used the mean weighted assemblage of the midway values from the node and its neighbors. One of the benefits of the score aggregation over node embedding is to trim down the overall number of total parameters. The initial reputation score r_x^0 of node x is measured using any neural network and feature vector $\vec{f}_x^0$ of the node. A fully connected ANN returned the node's initial importance based on network characteristics and knowledge, as shown in Eq.(5.6).

$$r_x^0 = \mathrm{ANN}(\vec{f}_x^0) \qquad \dots (5.6)$$

Further, the reputation aggregation score r_x^{η} of node x at layer η is calculated in Eq.(5.7).

$$r_x^{\eta} = \sum_{y \epsilon И(x) \cup (x)} r^{\eta-1}(y)\, T_{x,y}^{\eta} \qquad \dots (5.7)$$

where $T_{x,y}^{\eta}$ represents trust of node x on node y calculated by RA at η-th layer, $И(x)$ indicates the total number of first-order neighbors of node x, and $r^{\eta-1}(y)$ reputation aggregation score received from the neighbors in the preceding $(\eta - 1)$ layers. Thus in nutshell, The reputation aggregation score of a node x is represented as in Eq.(5.8)

$$r_x^L = \left\{\begin{matrix} \mathrm{ANN}\left(\vec{f}_C^0\right) & \text{if } L = 0 \\ \sum_{y \epsilon И(x) \cup (x)} r_x^{\eta-1}(y)\; T_{x,y}^{\eta} & \text{if } L > 0 \end{matrix}\right\} \qquad \dots (5.8)$$

The centrality reconciliation $d_{\eta}^{*}(x)$ is applied at the last concluding layer to include the centrality weightage of node x. Thus the concluding reputation score r_x^{*} of node x is the mean of the multiplication centrality reconciliation and reputation aggregation score of node x by using σ a lateral constant, as shown in Eq.(5.9).

$$r_x^* = \sigma(\text{mean}\,\{d_\eta^*\,(x) * r_x^L\}|\,\eta = 1,2,3,\dots,L) \qquad \dots (5.9)$$

5.3.2.4 Model training

It is required to train the model for measuring the reputation score accurately as much as possible to avoid overfitting and underfitting. Due to the social network's dynamic nature, it is very challenging to find the ground-truth about the network. Still, the GOLI model is trained by calculating the mean squared error between the obtained reputation score and the known reputation score t(i)from the set $\mathscr{R} \subseteq V$. Hence, the loss function S is represented using Eq.(5.10).

$$S = \frac{1}{V}\sum_{i \in \mathscr{R}}(r_i^* - t(i))^2 \qquad \dots (5.10)$$

5.3.2.5 Parameter setting

In this research, various parameters are used for calculating reputation score, trust, centrality, and model learning purpose. The weighted optimized value of these parameters is obtained during the model learning. The accurate parameter value enhances the predictive power and decreasing the training time of the model. Various issues are considered overfitting-underfitting, learning rate, batch size, and gradient descendants for local minima calculation. Thus, the following parameter values are chosen based on multiple tests, as shown in Table 5.2.

Table 5.2. Parameters value with the corresponding testing value range for GOLI model

Parameter	**Optimal value**	**Testing value range**
Learning rate(τ)	0.01	(0,0.5]
Batch size(B)	500	[100, 2000]
Total number of layers (L)	16	[2, 30]

Total number of epochs	400	[100, 1000]
$\vec{\delta}_\eta$	0.25	(0,1]
β_η, γ_η	Varies in each layer	(0,1]

5.3.3 Model's virtues

- **Sustain parallelism:** One of the main reasons behind choosing the GNN is that it maintains workflow parallelism, i.e., all the nodes perform the computation concurrently without any external inference. The calculation involves feature aggregation, reputation score aggregation, trust calculation, etc. Even during the model's training, all the nodes parallelly receive and send the information to other neighbors.
- **Capable of handling big data:** Often, most models and approaches can not handle the enormous data size, but the GOLI model is GNN based model that quickly takes a large amount of data. The dataset's density is a significant factor in finding the number of strong-weak ties and triangles. GOLI analyzed the latent features of nodes efficiently because of the profoundly layered architecture. As soon as the network's volume gradually enhances, all the nodes functioned efficiently and simultaneously aggregated neighbor's information.
- **Support a variety of datasets:** One of the most significant issues with the previous models is supporting a limited number of datasets applicable only in a finite domain. The proposed model is very efficient in holding a wide range of datasets. For authorization, we have used the diverse province's six datasets as input in the model and successfully capture the top-n opinion leaders for each dataset. Thus, it is applicable in various real-world applications.
- **Control network isomorphism:** The proposed model's leading merit is to manage the isomorphic network dilemma by converting the network isomorphic to graph isomorphic. In the network isomorphism, the two networks' structure looks different, but actually, they are identical and share the same features. Some of the models

produced different results for the isomorphic network, but the proposed model had the same effect for all isomorphic datasets.

- **Highly accurate results:** The proposed model successfully attained around 92% accuracy, 95.4% precision, 96% recall, and 95% F1-score for all the experimental over other standard SNA measures. Although some other models also produced highly accurate results, but only in the specific field. Our model covers a diverse range of datasets and lucratively achieved the desired target.

5.4 Experimental setup and results

In this research, six real datasets are used that represent different types of online social networks. The main issue behind selecting these datasets is to cover all kinds of activities and behavior happen on the social network. All the datasets have their features and services that support different types of information exchange. The graph analyzing software Gephi 0.9.2 [12] is explored for the revelation and investigation of network properties. Intel Core i7-9700K 9th Generation 3.6 GHz processor, python 3.6, networkx, PyTorch, and Tensorflow (with GPU momentum) libraries [186] are used for instruction interpretation as well as execution.

5.4.1 Datasets

5.4.1.1 Slashdot dataset

Slashdot datasets is an advanced technology-rich news broadcasting website in which website authors submitted their newsletters and reports related to advance technologies and techniques. Further, the editors are to evaluate these reports that made their decision about acceptance or rejection. One of Slashdot dataset features is tagging done by one author to other authors as acquaintance or rival. Therefore, an acquaintance or rival connection is established among the users. The dataset includes a total of 77360 users and 905468 edges among them [205].

5.4.1.2 Epinions dataset

Epinion dataset defines the 'who-trust-whom' network based on the customer comments on a particular product through the website' epinion.com.' So, first, the user has to make registration on the website and then able to submit their reviews on products. It merely depends on other consumers to follow other reviews or not. The user review then makes a web of trust and integrates the reviews rating to determine the customers' recommendation. The dataset has 75879 users and 508837 connections based on the user's suggestions [206].

5.4.1.3 LiveJournal dataset

LiveJournal is a famous social networking website for promoting and creating a diary, journal, or blog. If one user is not present in another user's friend list and has no mutual friends; still, they can mutually become friends without any contraction. The dataset consists of an overall 3,017,286 users and 87,037,567 directed links among them [207].

5.4.1.4 Last.fm dataset

Last.fm is one of the websites related to music and the user's recommendation. A user can seek a different kind of music according to their taste, choice, and advice. So, the website offers a flowless radio service and a variety of styles of music. The website also maintained a specific user profile that is the base for music suggestions. Last.fm network consists of a total of 136,420 users and 1,685,524 connections among them [208].

5.4.1.5 Bitcoin alpha trust weighted signed network dataset

Bitcoin alpha trust weighted signed network represents the trusted network in which various customers do the trading for Bitcoins using the portal called Bitcoin Alpha. In this trading, it is required to maintain the user's details and respective transactions as all the users' status are unidentified, i.e., users have a masked identity. To prevent forgery and supervise trustworthiness, all the members graded in the range of -10 to +10. The user with high positive grading has more reliability and vice-versa. There are overall 3783 nodes and 24186 edges in the network [188].

5.4.1.6 Weibo-Net-Tweet dataset

Weibo-Net-Tweet network pursues the 'who follows whom' kind of structure. Initially, some random number of users picked up from the network and then collected the total number of followers' followers. Recursively the same process has been carried out, and the collective number of large users formed a vast network. It is observed that, on average, every user has around 200 followers. The dataset includes approximately 1,776,950 users, and 308,489,739 edges exist based on user comments and retweets. We used the user's and their followers' profile information for analysis purposes [209].
Thus, The summarised description of each dataset is shown in Table 5.3.

Table 5.3. Summarized datasets description used for GOLI model

Dataset	Nodes	Edges	Density	Avg. clustering coefficient	Diameter
Slashdot	77360	905468	0.00032	0.0603	11
Epinions	75879	508837	0.00017	0.0578	14
LiveJournal	3,017,286	87,037,567	1.91*10-5	0.0371	16
Last.fm	136,420	1,685,524	0.00018	0.0822	8
Bitcoin alpha	3783	24186	0.00338	0.0523	7
Weibo-Net-Tweet	1,776,950	308,489,739	0.00019	0.0783	22

5.4.2 Estimated outcomes analysis and visualization

This section publicized the outcomes produced by applying the GOLI on the real data sets, as mentioned in the preceding section. Initially, each node's feature vector is measured based on multiple characteristics depending on network topology and neighbor's knowledge. It is critical to measure trust in the social network realm due to the users' unstable relations. So, few random values are consigned between 0 and 1 as a trust to achieve each node's experimental motive. Next, each node's initial reputation score is measured and passed the same information to the respective node's neighbors. Various parameters are used during the model training and efficient execution. Around 100 tests have been conducted to find the best possible value of parameters. The model's learning rate is 0.01, and 500 observations are put together in a single batch. A total of 16 layers or channels are used for network configuration and reputation transmission, while the entire process required around 400 epochs.

Further, the node's reputation score and trust are recursively forwarded to the next level layers, where RA computed each node's aggregated reputation score. Finally, the node with a higher reputation score is declared as the opinion leader. This research has applied the GOLI model on six different datasets; Epinions, LiveJournal, Last.fm, Bitcoin alpha, and Weibo-Net-Tweet, respectively. In this research also the value of five standard SNA centralities are calculated for the comparison purpose. Table 5.4, Table 5.5, Table 5.6, Table 5.7, Table 5.8, and Table 5.9 show the top-n(=5) opinion leaders in each SNA measure along with the used experimental dataset correspondingly.

Table 5.4. Top-5 opinion leaders for Slashdot dataset using GOLI model

User id	DC	User id	CC	User id	BC	User id	EC	User id	PR	User id	GOLI Reputation
23845	0.2344314	65402	0.1927551	26512	0.1447860	29087	0.1334619	40067	0.1145022	**39089**	**0.3022736**
65213	0.2344312	28712	0.1927550	39821	0.1447860	11090	0.13345618	3267	0.1145022	**71092**	**0.3022735**
40781	0.2344311	8956	0.1927549	6982	0.1447859	61890	0.13345618	18590	0.1145022	**67845**	**0.3022734**
1278	0.2344311	44231	0.1927548	55782	0.1447859	39055	0.13345617	72334	0.1145021	**56119**	**0.3022734**

55402	0.2344310	71092	0.1927547	1095	0.1447858	4902	0.13345616	50893	0.1145020	**26512**	**0.3022733**

Table 5.5. Top-5 opinion leaders for Epinions dataset using GOLI model

User id	DC	User Id	CC	User id	BC	User id	EC	User id	PR	User id	GOLI Reputation
27852	0.3487869	47720	0.2842291	52358	0.3047477	21158	0.1640095	7364	0.1765642	**20085**	**0.3573947**
19983	0.3487868	719584	0.2842290	39068	0.3047475	38694	0.1640093	43626	0.1765641	**37756**	**0.3573945**
345	0.3487868	20085	0.2842288	63954	0.3047474	8634	0.1640092	27575	0.1765640	**58671**	**0.3573944**
69434	0.3487867	639	0.2842287	28768	0.3047472	28468	0.1640090	58363	0.1765638	**19686**	**0.3573942**
54764	0.3487866	11593	0.2842285	13869	0.3047471	67364	0.1640089	11675	0.1765636	**7665**	**0.3573941**

Table 5.6. Top-5 opinion leaders for LiveJournal dataset using GOLI model

User Id	DC	User id	CC	User id	BC	User id	EC	User id	PR	User id	GOLI Reputation
2085943	0.1045491	675749	0.0985694	583920	0.1300784	860509	0.0767853	1185760	0.0938963	**759430**	**0.1389832**
29581	0.1045491	2960049	0.0985694	1956950	0.1300784	457302	0.0767852	748495	0.0938962	**1946397**	**0.1389832**
698043	0.1045490	57584	0.0985693	438596	0.1300784	295869	0.0767852	2986090	0.0938962	**597940**	**0.1389832**
584953	0.1045490	957482	0.0985693	857469	0.1300783	6859	0.0767851	48493	0.0938962	**28575**	**0.1389831**
1773829	0.1045489	1859684	0.0985693	29586	0.1300783	986849	0.0767851	986893	0.0938961	**699301**	**0.1389831**

Table 5.7. Top-5 opinion leaders for Last.fm dataset using GOLI model

User id	DC	User id	CC	User id	BC	User id	EC	User id	PR	User id	GOLI Reputation
59293	0.2933422	82364	0.1743229	85023	0.1699493	57232	0.1833385	28644	0.1418205	**9034**	**0.2744035**
107821	0.2933422	18432	0.1743227	64172	0.1699492	6693	0.1833384	65733	0.1418204	**29422**	**0.2744034**
8632	0.2933421	84561	0.1743226	448	0.1699492	28319	0.1833384	38376	0.1418204	**14880**	**0.2744034**
17901	0.2933420	117281	0.1743226	16482	0.1699491	118473	0.1833383	9251	0.1418203	**847**	**0.2744033**
14873	0.2933419	1167	0.1743225	59072	0.1699490	928	0.1833381	13932	0.1418202	**100982**	**0.2744032**

Table 5.8. Top-5 opinion leaders for Bitcoin alpha dataset using GOLI model

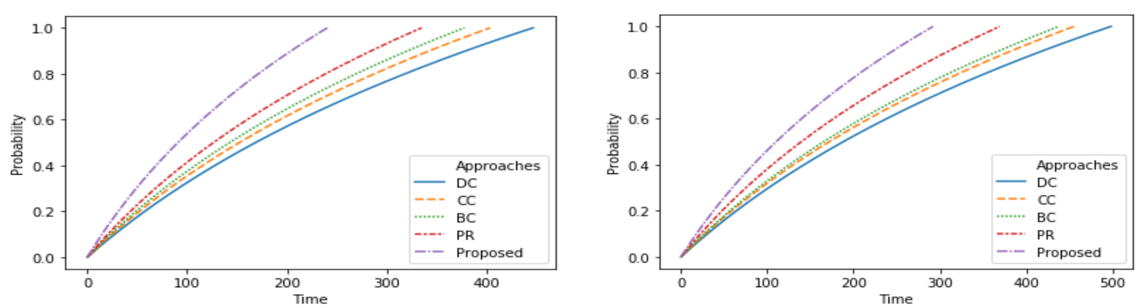
User id	DC	User id	CC	User id	BC	User id	EC	User id	PR	User id	GOLI Reputation
748	0.2178362	2083	0.1610389	839	0.1429575	920	0.1857445	476	0.192294	**2690**	**0.2289953**

1063	0.2178361	185	0.1610388	43	0.1429574	582	0.1857414	805	0.1922930	**698**	**0.2289952**
93	0.2178360	449	0.1610387	2782	0.1429573	2558	0.1857412	2639	0.192291	**1503**	**0.2289952**
384	0.2178359	63	0.1610387	736	0.1429571	1005	0.1857412	382	0.1922899	**812**	**0.2289951**
2559	0.2178358	1996	0.1610386	527	0.1429570	2026	0.1857411	1063	0.1922899	**448**	**0.2289950**

Table 5.9. Top-5 opinion leaders for Weibo-Net-Tweet dataset using GOLI model

User id	DC	User id	CC	User id	BC	User id	EC	User id	PR	User id	GOLI Reputation
94534	0.1223454	39984	0.1445641	58254	0.1840643	1159002	0.1334523	47093	0.1006747	**843514**	**0.2035639**
1403515	0.1223454	1045342	0.1445641	160036	0.1840643	263389	0.1334522	262363	0.1006746	**5380**	**0.2035639**
843514	0.1223453	564509	0.1445640	483284	0.1840642	724373	0.1334522	504993	0.1006746	**1489402**	**0.2035638**
201633	0.1223452	197846	0.1445640	35742	0.1840642	2903	0.1334521	925381	0.1006745	**77381**	**0.2035637**
583862	0.1223452	1347568	0.1445639	765627	0.1840641	534490	0.1334520	1205994	0.1006745	**1026185**	**0.2035637**

Further, in the social network, the user does not only portray like a recipient and the activity and behaviors play a major role in the evolution of network and information diffusion. Thus, one of the critical chores with the social network is how to measure the diffusion of information. A probability distribution function p used to measure the number of users influenced over time. In Fig. 5.6, we have illustrated the information dissemination process where the x-axis represents the time (in sec.) while the y-axis depicts the probability distribution function.



(a) (b)

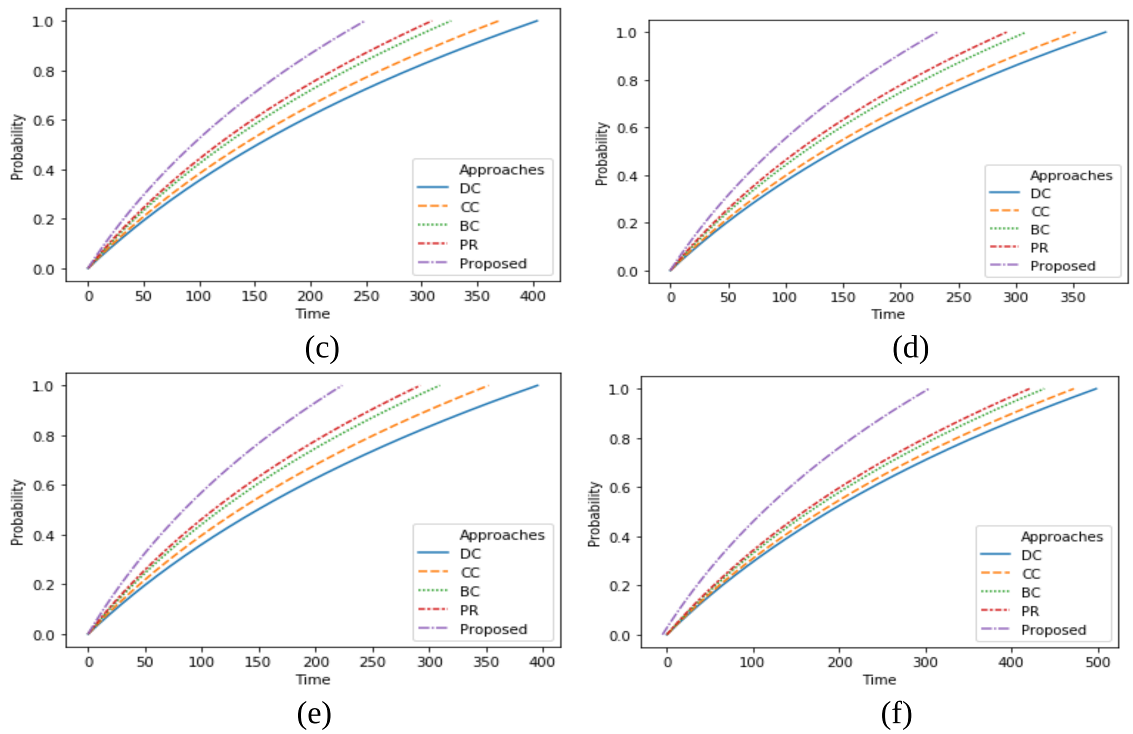


Fig. 5.6. Visualization of Information diffusion using GOLI model over time for (a) Slashdot, (b) Epinions, (c) LiveJournal, (d) Last.fm, (e) Bitcoin alpha, and (f) Weibo-Net-Tweet dataset

For instance, if the value of p = 1, i.e., all the nodes get affected, and information has been spread in the entire network. It is observed that the GOLI model needed lesser time for information dissemination over other SNA measures for all the datasets.

5.5 Performance evaluation

A variety of parameters are used to boost and scrutinize the performance of the GOLI model. To observe the model performance, essential performance evaluators put together data splitting, training accuracy, testing accuracy, and error percentage into consideration. Often, the best parameter value also affects the model's performance by overcoming the overfitting or underfitting problem [210]. So, approximately 60% of the total data used for training, 20% for validation, and 20% for testing operations. In Fig. 5.7, visualization of the GOLI model's training and testing accuracy for all six datasets is demonstrated. It is found that GOLI obtained around 91% training accuracy and 92% testing accuracy with an approximately 1% error rate.

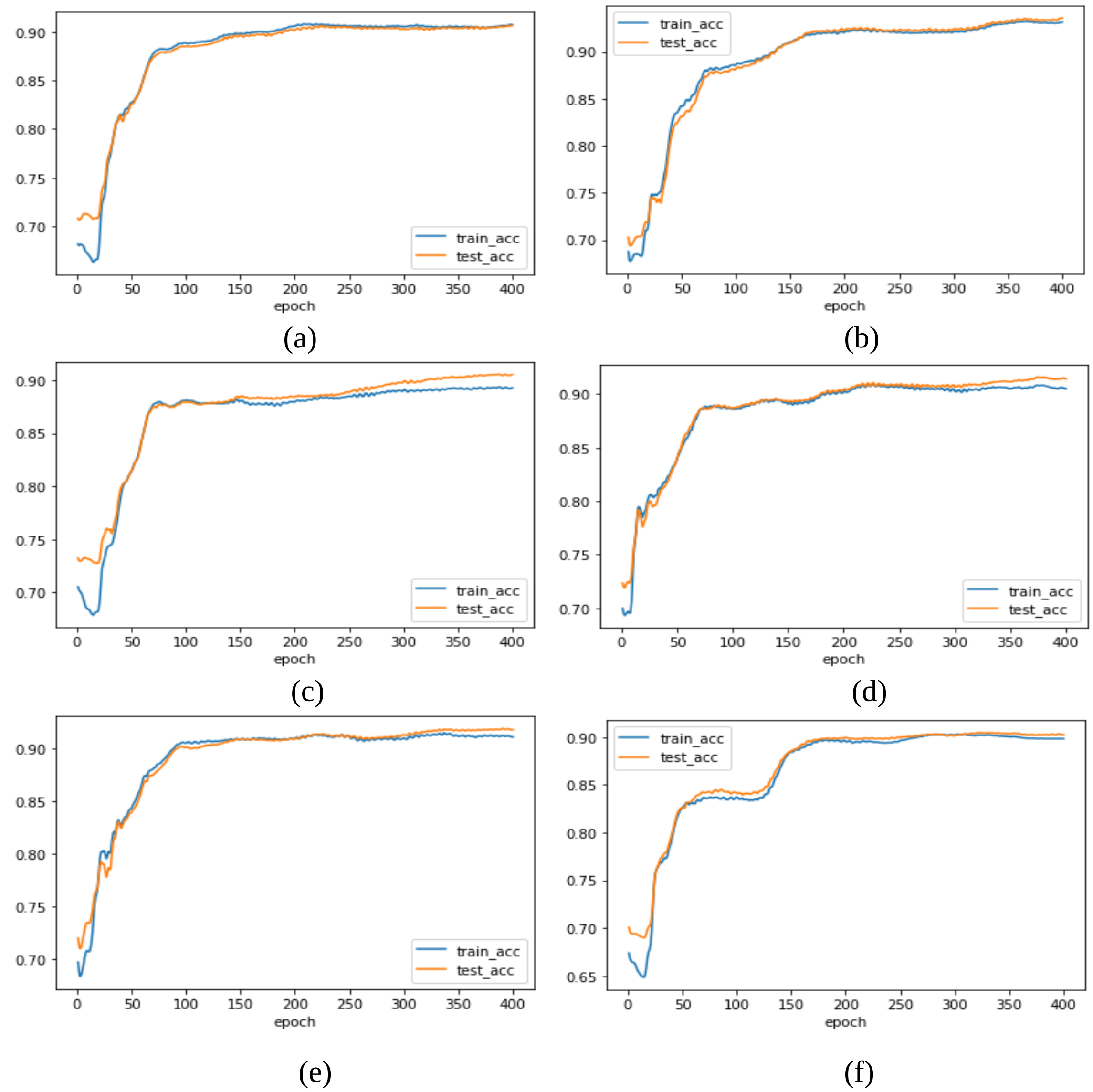


Fig. 5.7. Training and Testing accuracy visualization for (a)Slashdot, (b)Epinions, (c) LiveJournal, (d) Last.fm, (e) Bitcoin alpha, and (f) Weibo-Net-Tweet dataset

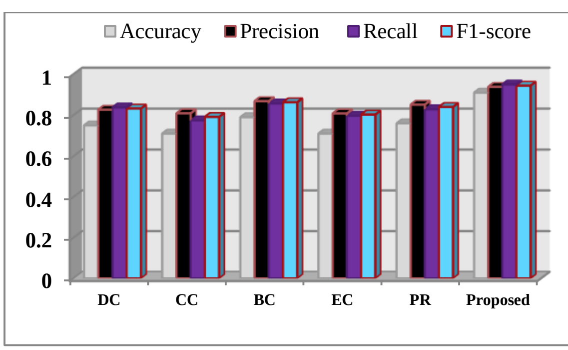
In Table 5.10, the discrete value of training accuracy, test accuracy, and the error rate is mentioned for all the experimental datasets.

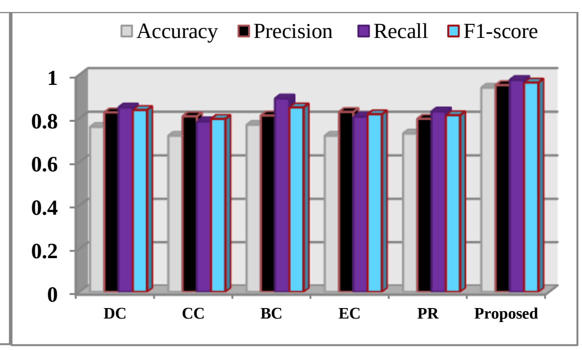
Table 5.10. Training accuracy, Test accuracy, and Error rate for all the datasets

Dataset	Training accuracy	Test accuracy	Error rate
Slashdot	0.91684 ± 0.00111	0.91881 ± 0.00152	± 0.0226%
Epinions	0.94738 ± 0.00019	0.94729 ± 0.00104	± 0.0027%
LiveJournal	0.90841 ± 0.00273	0.92267 ± 0.00084	± 2.0244%
Last.fm	0.91328 ± 0.00122	0.92027 ± 0.00042	± 1.0028%
Bitcoin alpha	0.91454 ± 0.00104	0.92829 ± 0.00128	± 1.0819%
Weibo-Net-Tweet	0.90492 ± 0.00016	0.91205 ± 0.00057	± 1.0025%

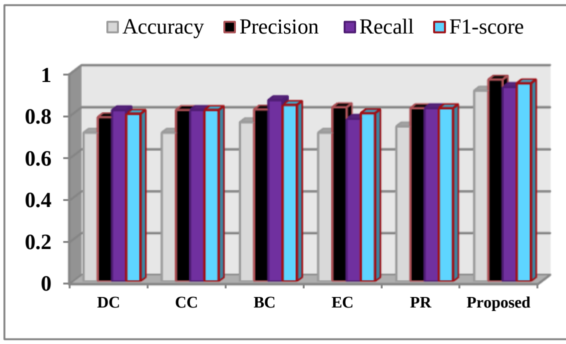
5.5.1 Performance metrics

Various metrics are used to check the quality, productivity, and efficiency of the model. The grounded and most straightforward method is to calculate the Accuracy, Precision, Recall, and F1-score of the model [211], [212]. Next, based on the network's pre-defined ground truth information, the progressive comparative outcomes are found, as shown in Fig. 5.8. It is also observed that the GOLI model offered improved qualitative results over other SNA measures for all six datasets. The addressed GOLI model produced approximately 92% accuracy, 95.4% precision, 96% recall, and 95% F1-score over other SNA measures.

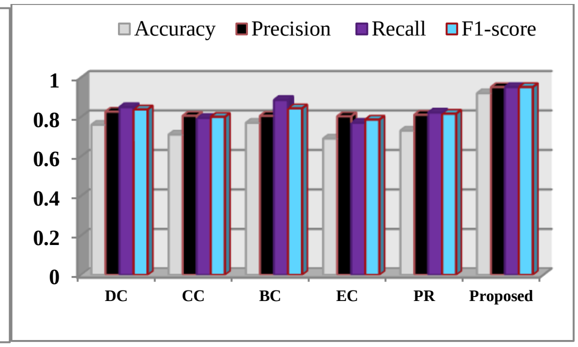




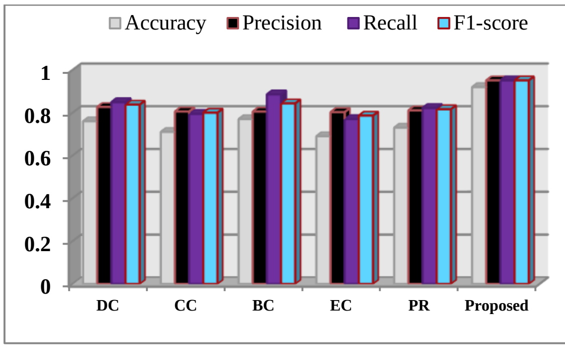
(a) (b)



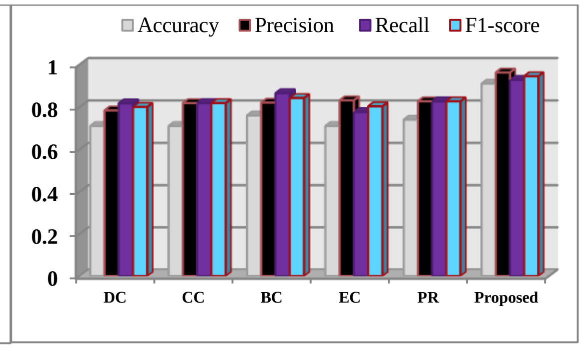
(c)



(d)



(e)



(f)

Fig. 5.8. Performance metrics contrastive analysis between standard SNA measures and (a)Slashdot, (b)Epinions, (c) LiveJournal, (d) Last.fm, (e) Bitcoin alpha, and (f) Weibo-Net-Tweet dataset

Another issue that also impacts a model's performance and plays a vital role in the model's selection is time complexity. The model's training required a certain amount of time to gain knowledge about the nature of the network. The entire computation time included data preprocessing, data loading, model training, code execution, and result visualization. The whole execution time required for the other standard SNA measures and GOLI model is demonstrated and compared in Fig. 5.9.

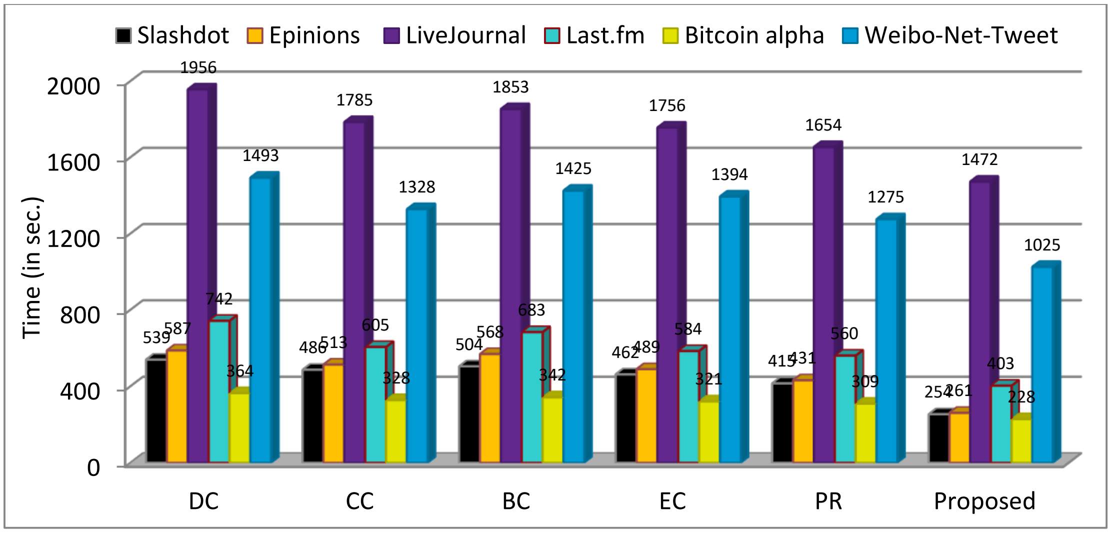


Fig. 5.9. Time complexity comparison between standard SNA measures and (a)Slashdot, (b)Epinions, (c) LiveJournal, (d) Last.fm, (e) Bitcoin alpha, and (f) Weibo-Net-Tweet dataset

5.6 Chapter summary

In this chapter, the power of GNN is used to design the GOLI model for measuring the user's reputation in the network. Initially, the latent features of the nodes measured and constructed the feature vector matrix. Further, the information is transmitted and aggregated by the neighbors to compute the node's reputation score. Based on the reputation score and trust, an anticipated trust network formed to fetch the network's trustworthiness. Next, at every layer of the model, reputation score and trust are aggregated as per the proposed method based on neighbor's information. Finally, to include the importance of in-degree centrality, a centrality settlement formulates to calculate the node's final reputation score. To end with, based on the highest reputation score, a list of user prepared and top-n(n=5) users declared as opinion leaders. The findings are compared with the other standard SNA measures for six real-world datasets to evaluate the model's strength. It is analyzed that the proposed model produced improved results w.r.t. evaluation metrics (accuracy, precision, recall, F1-score, error rate, execution time) and achieved the precise opinion leaders.

The applicability of the GOLI model is too significant for identifying critical nodes in various realm graph-related problems. Traditionally, opinion leaders' scope includes a broader range of domains such as online campaigns, advertising, marketing, healthcare, agriculture, pooling, recommendation system, finance, defense system, etc. As soon as the era and technologies change, opinion leaders' role and accountability also revolutionize significantly. The GOLI model's pertinence would also be appreciably helpful in the diverse other modern applications such as catalyst analysis in the chemical reaction, GIS, biological pattern, modern algebra, regression analysis, cloud computing, load balancing, distributed computing, and many more.

The proposed model also has some limitations. The first restraint with the model is that it works on the static dataset, so we would like to experiment on the proposed model with a dynamic dataset considering user activities over time [213], [214]. As technology evolves, the new GNN-based model also comes into view for graph-based data. Therefore, the variant of GNN with the online social network would be further impactful and exciting for perfect opinion leader detection. Besides, we would also explore the more advanced deep-learning-based models associated with other user's multi-relational characteristics such as user's response time, geographical location, the domain of interest, etc., as upcoming directives.

Chapter 6

Applicability of opinion leaders to control COVID-19 rumors in online social networks

Since the last few time, the COVID-19 pandemic has spread across the world swiftly at an alarming rate. The disease is being considered a global crisis from the health and economic view. In this chapter, opinion leaders' applicability is introduced by controlling the COVID-19 rumors that pull human physical and mental health into risk. Section 6.1 covers the overview of the chapter. The entire proposed methodology has explained in section 6.2. A Reputation-based Opinion Leader Identification (ROLI) algorithm is also defined to discover the opinion leaders in the Twitter dataset. The complexity of the approach is discussed in section 6.3. Section 6.4 demonstrates the information about the datasets and experimental outcomes. The complete summary of the chapter is wrapped up in Section 6.5.

6.1 Overview

On February 15, 2020, World Health Organization director-General Tedros Adhanom Ghebreyesus stated, "We're not just fighting an epidemic; We're fighting an infodemic," about COVID-19 in the virtual Munich Security Conference [215]–[217]. The term "infodemic" means to face complications to find out the answer to a problem due to an excess quantity of propaganda and rumors spread all over online and offline. So, the rumors lead to misinformation about the thing and produce mental fear, distress, and social disorder [218]. It deliberately makes agonize about the density of the COVID-19. In social media, COVID-19 rumors spread very rapidly without any authentication and legal verification. Even a single bit of illegal rumor outbreaks the entire society badly and instigates health consciousness [219], [220]. Eventually, WHO came forward has decided to make their myth-buster that clarifies and diffuses most of the myth about the COVID-19 [221]. It includes various public advice, videos, and preventive measures essential to control the disease. The various aspects that influence an individual to transmit the rumor are discussed, as shown in Fig. 6.1.

Fig. 6.1. Factors influence COVID-19 rumors

- **Informative:** If a particular rumor contains some new information regardless of its veracity, people think to spread that rumor rapidly to gain knowledge about the information. In the case of COVID-19, most of the rumors spread the cause of innovative concepts related to vaccine and avoidance.
- **Biasing towards belief:** According to the research, people are relatively more partial towards their beliefs and values. They can readily admit and reciprocate the rumors generated by their nearby surroundings, family groups, and workgroups. So this is also the leading factor for the spreading of COVID-19.
- **Trusted sources:** If multiple persons belong to the common community and received COVID-19 rumors generated from an outsider trusted source, it is likely to be possible that the same rumor is also be accepted or disseminated by the other members of the community. So it enhances the probability to spread the rumor fast.
- **Self-image:** most of the time, people want to create their self-image among others in their social circle. Thus, due to this misconception, some COVID-19 rumors are drastically spread over social media. Once the people receive any rumor, they simply disseminate it and think that it would extend their self-image in their workgroup.

- **Relief from panic:** In some situations, it has happened that people spread the rumors so that others can take the befit of information embed in it because each rumor does not have misinformation. Some rumors about COVID-19 also have a small amount of information that can partially avoid the possibility of COVID-19. For example, lukewarm water or salted water can control the virus that causes COVID-19, but it can not kill them.
- **Social status:** As per human psychology, If a person spreads any information as soon as received, it will enhance the social position in society. So this misconception is one of the root causes of most of the rumor transmission. Similarly, in COVID-19, people forwarded different types of messages and posts on various social media without checking their veracity.
- **Uncertainty:** When people are uncertain about the rumor's integrity, they just spread it to know its trustworthiness. In the early stage, people were not sure about the symptoms and preventive measures of COVID-19, so they just forwarded it as they received it from another group through WhatsApp, Facebook, Twitter, etc.

Social media plays a critical role in widespread public events and activities. Sometimes they provide the white side of the picture and sometimes unknowingly or deliberately promote an image's dark side [220]. So it is social media's rationale to decide whether to filter out the complete information or make it viral. They also provide the facility to access a vast amount of information to achieve a unified goal [222]. Whenever a specific emergency happens in the real world, different tweets, posts, and messages are moved around social media without knowing the integrity [223]. Even though the national government agencies are ready to face the crisis and support providing all necessary services, all such facilities are brought down and in suspicion because of various false rumors. The main reason behind the spreading of the rumor is blind trust without knowing the facts. Rumors spreading in those communities are very rapid. They align with the previously presented values, i.e., people already have trust to a certain degree in the community's other populace. So, when the people struggle for basic needs and the existing condition is crucial, such a situation stimulates the approval of rumors.

According to the sources, thousands of deaths occurred due to rumors and fake information in the universe's different regions. In various countries, numerous research and study groups are developed to identify the rumors spreading rate, misinformation, types of rumors, and measuring the impact of rumors during COVID-19 [224]. Some of the researchers seemed that SARS-CoV-2 is accountable for the spreading of COVID-19. Few people stated that COVID-19 is sent out through the 5 G network, and because of this rumor, a few of the 5 G towers were also destroyed by the people in some countries. Further, various rumors spread over social media regarding the different medical treatments such as alcohol, garlic, ginger, warm water, etc., that can help control COVID-19. Nowadays, many rumors related to the COVID-19 vaccine are floating on social media [225], [226]. In a nutshell, some rumors have found that they are too active on social media as follow:

- Garlic and high temp water both are proficient protectors.
- Pneumonia immunizations/anti-infection agents help to secure.
- It affects just more seasoned individuals; kids are insusceptible.
- High temperature obliterates the infection.
- Protection through Drinking liquor.
- Thermal scanners are helpful to detect.
- Pepper essence in the food can prevent COVID-19 infection.
- Houseflies transmitted the COVID-19 virus.
- Any kind of bleach or decontaminator on the body defends against COVID-19.
- The mobile signals (5-G) transmitted COVID-19.
- The sun heat or higher temperature guard from COVID-19.
- If a person can hold the breath for ten or more seconds without any uneasiness, it means he/she is not affected by COVID-19.
- Taking a hot shower will free a person from COVID-19.
- The COVID-19 virus can be killed by the use of hand dryers.
- Vaccines used for the treatment of pneumonia prevent the COVID-19 virus.
- Paracetamol or any other antibiotics can cure COVID-19.

Thus, a practical method is needed to control the social damage from the COVID-19 rumors and spread only likely to be positive information. During this research, it is observed that users have posted various rumors and misinformation related to COVID-19 globally. In Fig. 6.2, the list of top-15 countries has displayed whose users have placed maximum numbers of Twitter posts. United States of America (USA) secured the top position with 51.61% of the total tweets posted while the United Kingdom (UK) and India clutch the second and third rank, respectively.

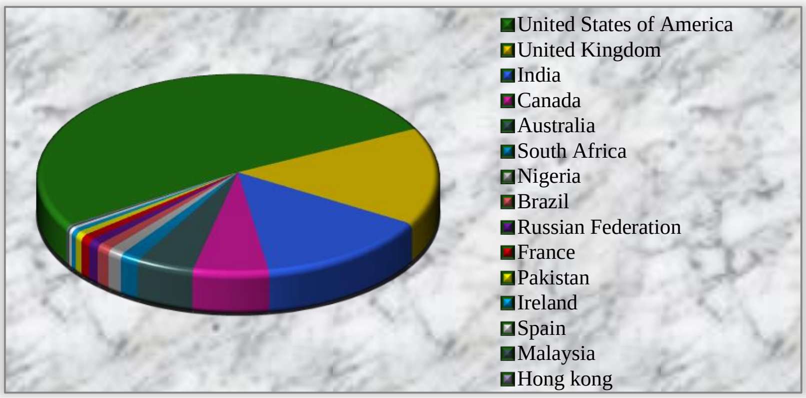


Fig. 6.2. List of Top-15 countries' users posted the maximum number of COVID-19 tweets

The power of opinion leaders can modify the perception and awareness about the severity of COVID-19. They rigorously examine the integrity by measuring the source information and inspect the authenticity of the rumor. It is also essential to scrutinize opinion leaders' influence and information dissemination in the social network to triumph over COVID-19 and potential epidemic outbreaks.

In this researcher, the opinion leader behaves like a represser who control rumors from their skills and knowledge. Opinion leaders substantially impact and prevent their supporters and believers from any risk [55], [227]. So, suppose they follow any pattern or advice not to promote a particular object or material for society's benefit. In that case, their followers will most probably follow the same guidelines as shown in Fig. 6.3. When any users interact with either opinion leaders or rumor spreaders, it may be possible to change their viewpoint or consistent with their beliefs.

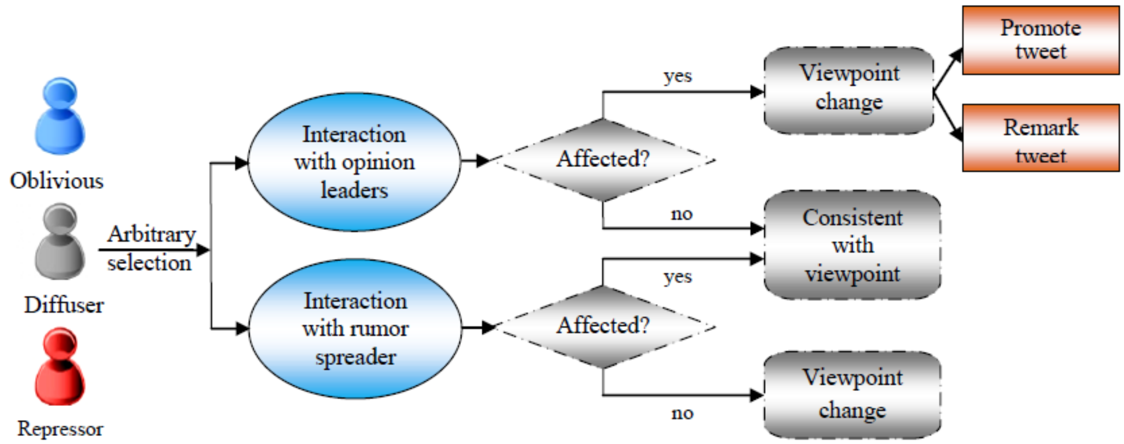


Fig. 6.3. Human influenced viewpoint representation

Thus, a unique approach is developed that illustrates opinion leaders' power by controlling the COVID-19 rumors as much as possible. Initially, a large number of tweets has been extracted from Twitter social network related to COVID-19. Next, all the tweets are preprocessed by handling negations, punctuations, degree modifiers, word-shapes, emoticons, emojis, slang words, and sentiments-laden. Further, the reputation of each node is measured to figure out trust. The paramount significance of trust is to verify the trueness of the tweet so that if any user spread any information on the social network, the opinion leader can confirm the tweet's authenticity [143]. A ROLI algorithm is used to target the list of top-T opinion leaders who are responsible for controlling rumors [228]. Finally, each tweet's entropy is measured based on the probability that the other user responds to the tweets along with estimated trust. If the tweet's entropy is less than the pre-defined threshold value λ, it seems to a rumor, and a report is generated against it; otherwise, it has been transmitted in the network.

6.2 Proposed methodology

In the proposed system, three types of nodes are defined- oblivious node, diffuser node, and represser node that is acting as the main character in the network. When an individual establishes a connection with a diffuser user, a rumor is transmitted from the diffuser to another. So the diffuser node is a user who propagates the rumor from one person to another. They have a serial connection with the other users and involve strongly during rumor diffusion. An oblivious node is a user who has neutral behavior during the rumor spreading,i.e., they do not take any action on the rumors and neither stop nor forwarded them. The represser node is a user who never transmitted it to others whenever they received the rumor and tried to stop the rumor's transmission.

For example, consider the network, as shown in Fig. 6.4. There is three colors node; The grey color nodes indicate the oblivious node. A green node represents the diffuser nodes, while red nodes depict the represser nodes in the network. When the represser node influences the diffuser node by its intelligence, it becomes a represser node. In the same way, the fundamental knowledge about the rumor veracity transmitted in the network gradually. Sometimes, few nodes are not much influenced by the represser node and firmed on their decisions. Thus, it is very complicated to divert others' decisions through social media; Still, opinion leaders try to persuade others' attitudes through their expertise [55], [229].

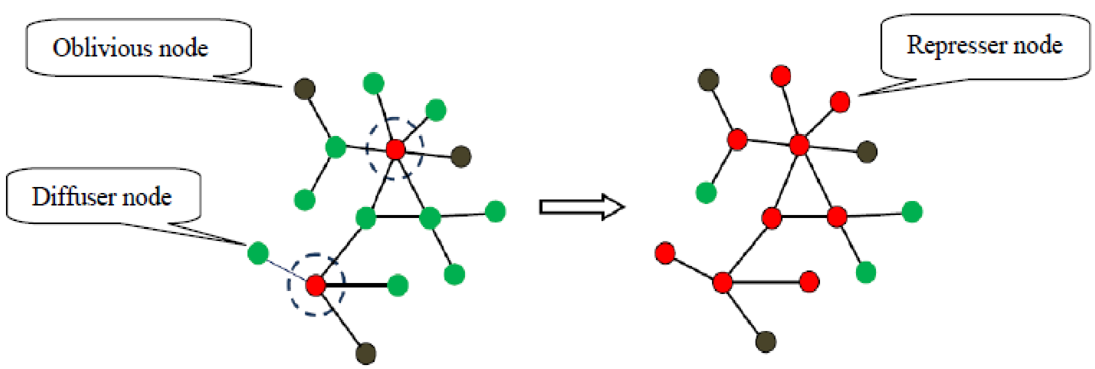


Fig. 6.4. Network representation with diffuser, represser, and oblivious node

Many tweets are initially fetched from social media to discover the population's polarity and sentiments effectively in this research. A VADER python library is used to find the tweet's polarity covering various features and opinions [230]. Further, each user's reputation is computed based on the polarity scores received from the neighbors and other nodes. Next, trust is measured to validate the trueness of the posts. Finally, the reputation-based Opinion leader identification algorithm is explored to find the top-T opinion leaders who check whether the posted tweet is a rumor or not. Thus, the proposed approach's overall structure and pseudocode are shown in Fig. 6.5 and algorithm 6.1, respectively.

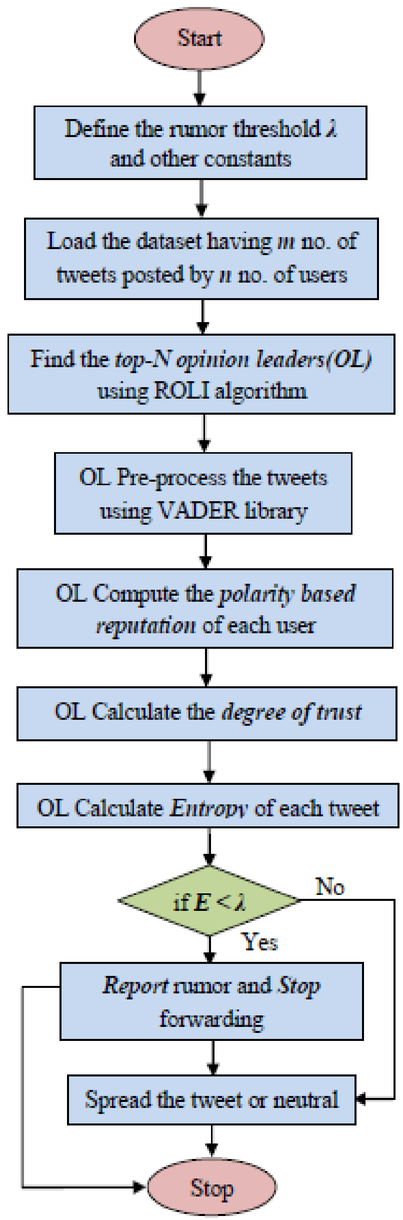


Fig. 6.5. Flow chart of proposed COVID-19 rumor controlling approach

---

Algorithm 6.1: Opinion **L**eader based **R**umor **D**etection **(OLRD) Algorithm**

Input:

1. Rumor threshold λ
2. Total m number of tweets posted by n number of users

Output: Decision about rumor spreading

Steps:

1. Apply the ROLI algorithm to find the top-T opinion leaders.
2. for $\forall$ t in m do

    Preprocess t by removing iterating characters, hashtags, hashtags, and URLs.

3. end for;
4. for $\forall$ j in n do

    Calculate the polarity-based reputation $r_j(t)$.

    Compute the degree of trust T_j.

5. end for;
6. for $\forall$ t in m do

    Measure the entropy $E_t(x_i)$ of each tweet.

    if ($E_t(x_i) < \lambda$)

        Report the tweet as a rumor and discontinue the rumor spreading.

    else

        Transmit the tweet normally.

7. end if;
8. end for;

6.2.1 Tweet Stage transition

6.2.1.1 Diffusion stage

In this stage, the diffuser user attempted to spread the rumor with the rate β in the network. The user finds the appropriate piece of information that is to be transmitted. Sometimes, the diffuser may modify some of the information according to the current circumstances and spreads the indistinct information to its neighbors. If the nodes do not want to participate in the transmission or having no facts, they simply ignore the transmission process.

6.2.1.2 Recognition stage

In this stage, the node decides whether to accept the rumor with the probability $q_d(t)$. If the degree of trust is very high on the other users, the node may accept the rumor and propagate it to its neighbors, but the node rejects the rumor if the degree of trust is low.

6.2.1.3 Probability of retweet

In the social network, millions of users posted various posts and retweets related to different events every second in a whole day with the rate of θ. Whenever the users refresh and reload the social media interface, the user finds various new posts. The user then decides to either read, forward, retweet, or ignore the posts. Let u_1 is the total number of first-order followers and u_2 is second-order followers of a user x. The second-order followers considered only those who retweet or read the posts read or retweet by first-order followers. Let, u_d indicates the total number of followers in the individual network, and $q_d(t)$ depicts the probability to find the total number of followers who pursue, read, and retweet the tweet at level d after its posting within time t. We have assumed that most users who read and also retweet the post are equal to p. Let $R_d(t)$ represents the total number of followers who read or retweet the tweets at level d after time t. Therefore, the value of $R_d(t)$ can be computed using Eq.(6.1).

$$R_d(t) = u_d(t)q_d(t)p \qquad 6.1)$$

In this process, the occurrence of a total number of tweets and the appearance of the total number of l-order users are considered exponential. So, poisson distribution is used to measure the total number of users who randomly read and retweet the posts at time t, represented using Eq.(6.2).

$$\int_0^t f(\tau)d\tau = \int_0^t \theta\, e^{-\theta t}\, d\tau = 1 - e^{-\theta t} \quad \dots (6.2)$$

So, the probability $q_1(t)$ for first-order users to retweet, can be represented using Eq.(6.3).

$$q_1(t) = 1 - e^{-\theta t} \quad \dots (6.3)$$

Now, we have formulated the Poisson distribution for second-order users. Here, one interesting fact is that the second-order users only retweet the post that is already retweet by the first-order users on or after time t. Let, τ considered the time on or before the first-order user read the tweets such that $\tau \leq t$. So, the probability $q_1(t-\tau)$ for first-order users to retweet, the post between t and τ can be depicted using Eq.(6.4).

$$q_1(t-\tau) = 1 - e^{-\theta(t-\tau)} \quad \dots (6.4)$$

So, the probability $q_2(t)$ for second-order users to retweet, can be represented using Eq.(6.5).

$$\begin{aligned} q_2(t) &= \int_0^t \theta\, e^{-\theta t}\, p q_1(t-\tau)d\tau \\ &= \int_0^t \theta\, e^{-\theta t}\, p\,(1 - e^{-\theta(t-\tau)})q_1(t-\tau)d\tau \\ &= p\,(1 - e^{-\theta t} - \theta t e^{-\theta t}) \quad \dots (6.5) \end{aligned}$$

Correspondingly, the probability $q_2(t-\tau)$ can be shown using Eq.(6.6).

$$q_2(t-\tau) = p\,(1 - e^{-\theta(t-\tau)} - \theta(t-\tau)e^{-\theta(t-\tau)}) \quad \dots (6.6)$$

Similarly, the probability $q_3(t)$ of retweet for third-order users is represented using Eq.(6.7).

$$\begin{aligned} q_3(t) &= \int_0^t \theta\, e^{-\theta t}\, p q_2(t-\tau)d\tau \\ &= \int_0^t \theta\, e^{-\theta t}\, p^2 \left(1 - e^{-\theta(t-\tau)} - \theta(t-\tau)e^{-\theta(t-\tau)}\right) d\tau \\ &= p^2\,(1 - e^{-\theta t} - \theta t e^{-\theta t} - \tfrac{1}{2}\theta^2 t^2 e^{-\theta t}) \quad \dots (6.7) \end{aligned}$$

Thus, the general formula to calculate the probability by the user at level-d is shown using Eq.(6.8).

$$q_d(t) = p^{d-1} \left(1 - \sum_{u=0}^{d-1} \left(\frac{(\theta t)^u e^{-\theta t}}{u!}\right)\right) \quad \dots (6.8)$$

If an individual received a bit of information from the other users in the network, it depends on the degree of trust that the user perceives from the other user. So, $q_d(t)$ measures the probability of reading and retweeting the post that the other users published in the network.

6.2.2 Tweet preprocessing and reputation calculation

Generally, a tweet is significantly affected by the active user's response, environment, content, time interval, and domain. The other's user tweet also influences the content and the schism of the tweet over time. So, we have redesigned our dataset in which each tweet is represented by the tuple <#user_id, content, time>. To measure each tweet's polarity, we have used the VADER (Valence Aware Dictionary and sEntiment Reasoner) python tool that categorized each tweet into the positive, negative, and neutral categories. This tool also defines the potency of positive, negative, or neutral polarity based on character case sensitivity, emojis, punctuation, emoticons, and slang. The primary elements of the tools include degree modifiers, conjunctions, and n-grams. For measuring the reputation of a user in the social network, we just used the modified reputation measuring course of action. Thus the reputation $r_x(t)$ of the node x is measured using Eq.(6.9), Eq.(6.10), and Eq.(6.11).

$$r_x(t) = \frac{1}{\partial}\sum_{i=1}^{x-1} \eta(r_i(t)) * r_{i+1}^{other} \quad \dots (6.9)$$

Where,

$$r_i(t) = \frac{\sum_{s=1}^{p}\sum_{m=1}^{q}\left(\sigma(p_{s,m} - n_{s,m}) + \tau u_{s,m}\right)}{\sum_{s=1}^{p}\sum_{m=1}^{q}\left(p_{s,m} + u_{s,m} + n_{s,m}\right)} \quad \dots (6.10)$$

$$\eta(r_i(t)) = 1 - \frac{1}{1+e^{-((r_i(t)-d)/\alpha))}} \quad \dots (6.11)$$

In the above equations, $p_{s,m}$ indicates the positive polarity of the m-th tweet, $n_{s,m}$ indicates the negative polarity of the m-th tweet, and $u_{s,m}$ indicates the neutral polarity of the m-th tweet. A constant σ is the weightage assigned to positive and negative polarity, and τ is another weightage constant allocated to neutral polarity. Again, α and ∂ are non-zero integer constants whose value lies between 0 and 1. $r_i(t)$ is the reputation of the user derived from a particular m-th tweet based on positive, negative, or neutral polarity, and

r_{i+1}^{other} is the reputation of other users who retweet on m-th tweet. $\eta(r_i(t))$ is a moderating function to guarantee that the reputation of the authenticated user becomes robust against any forged post at level-d. We have performed various tests to find the appropriate value of $\sigma,\ \partial,$ and τ. So, for experimental purposes, we have used $\sigma = 1,\ \partial = 0.85,$ and $\tau = 0.75$. We practically found that the impact of these parameters affects the reputation of a user in the network.

6.2.3 Trust computation

Trust plays an essential role in the rumor spreading over time. Trust depicts the belief that users gain from other users by their activities and actions in the network [140]. The role of trust is significant for accepting or rejecting rumors. The representation of trust is complicated as the variations in the user's preferences and attributes. Most users have thousands of friends on social media, but only a few hold the user's trust. Reputation portrays an essential character for trust computation [231]–[233]. If a user has a high reputation, most probably the user would trust them; But only in few cases is the user's reputation independent of the reputation. In the proposed approach, trust is calculated based on the reputation that a user achieves from neighbors and other users over time t in the network. We have computed the user y trust on user x by utilizing the user's reputation as shown in Eq.(6.12).

$$\mathrm{T_{x,y}} = \frac{\exp\left(\sigma\left(\sum\left(\mathrm{r_x}(t)|\times|\mathrm{r_y}(t)\right)\right)\right)}{\sum_{z\epsilon N(x)\cup\{x\}}\exp\left(\sigma\left(\sum\left(\mathrm{r_x}(t)|\times|\mathrm{r_y}(t)\right)\right)\right)} \qquad \ldots (6.12)$$

Where $\mathrm{r_x}$, $\mathrm{r_y}$, and $\mathrm{r_z}$ is the reputation score of user x, user y, and user z, respectively. N(x) is the set of neighbors of user x, σ is a normalization component with $0 < \sigma < 1$, and $|\times|$ is a multiplicative operation.

6.2.4 Tweet's entropy calculation

A tweet's entropy signifies the importance and amount of information that a tweet perceives during the transmission. Promotion and advertisement-like tweets are having less entropy, while news and informative-like tweets likely having more entropy. Generally, the user retweets or forwards only those tweets with some new and unique information or originated from some authorized source. There is a probability $q_d(t)$ to determine the chances of retweet on the post of user x by other users. So, the user x, i-th tweet overall entropy $E_t(x_i)$ at time t is calculated using Eq.(6.13).

$$E_t(x_i) = -\sum_{y=1}^{z} T_{xy} * (q_d(t) \log(q_d(t))) \quad \dots (6.13)$$

6.2.5 Opinion leader identification

The concept of the opinion leader identification algorithm came from product awareness in the field of product marketing. As a business strategy, the product manager's main motive is to promote the product by recognizing the group of users having a large number of followers on social media. The product managers try to attract only these information spreaders by providing them a few rewards or benefits. So it is a very classical problem to choose such a spreader that can maximize the total selling of the product regardless of network topology and other competitive business strategies. Hense such types of users are considered opinion leaders who have the power to control other nodes. Generally, most opinion leader identification approaches are based on degree centrality, closeness centrality, betweenness centrality, PageRank, and eigenvector centrality. Since the last few years, The PageRank-based procedure provides better outcomes in most cases and has received more deliberation towards solving the significant user identification problems.

A new Reputation-based Opinion Leader Identification (ROLI) algorithm is proposed that identified the most important opinion leader in the network based on the highest reputation. A higher reputation in the network has more chances to get more votes from its neighbors. So, identifying the user's reputation is very critical and depends on the total no of tweets and retweets posted by the user's neighbors. The ROLI algorithm's general concept originated from the SIR epidemic model in which a user exists in any of the three stages: susceptible, infected, and recovered. A variety of researchers have observed that if the infection rate of disease is high compared to the recovery rate, it is complicated to eliminate the illness that formed a pandemic. Every time an infected node attempted to spread the disease with the rate β to any of its neighbors.

Similarly, a node can be recuperated with the rate γ over time. Hence, the ROLI algorithm is used in which a node voted to its neighbors based on their reputation score. Initially, each node's reputation is calculated based on their total degree, i.e., the absolute number of nodes. Most of the social networks follow the richer gets richer phenomenon, i.e., the most potent nodes tending to magnetize more other nodes in the network. So, in the real world, if a node is having a higher level of trust or weighted relationship with its neighbor or other nodes, they would prefer to vote for that node. In this algorithm, there is a need to choose the top-T opinion leaders; every player has a chance to vote t times. If a particular user is selected as an opinion leader in one round, that node will not participate in further voting rounds. This strategy's main reason is to avoid biasing among the nodes because the spreading power of the elected node may influence or affect other nodes' decisions. So as the elected node would not be involved in voting, the subsequent node's neighbors and their neighbor's power also shrink. Hence the selected nodes are separated from the entire process and can not use their control unnecessarily with the neighbors. Finally, after t number of rounds, top-T opinion leaders have been selected successfully.

In this algorithm, a value pair (vs_x, v_x) is associated with every node, where vs_x depicts the voting score received from the node's neighbors and v_x indicates the node's voting ability, i.e., the total number of votes that the node can grant to its neighbors. Initially, all the nodes with the same capacity in the first round, i.e., one, to vote. Each node can offer a vote to its neighbors, and all neighbors can also give the option to the subsequent node. The voting score of the node is the collective aggregation of the votes that its neighbors have given, i.e., if a node receiving a total of five votes from its neighbors based on its reputation, the voting score of the node would be five. After each round, the node with the highest voting score would be declared as the opinion leader. It is also noticed that the elected opinion leader would not participate and set its voting score to zero in the subsequent round. All the nodes connected with the previous round's opinion leader have to update voting capacity in the following phase. The node update its voting capacity by

$(v_x - f)$ until the value of v_x reach to zero. In this update mechanism, f indicates the diminishing variable whose value lies between 0 and 1. For simplicity, the variable f is measured as $\frac{1}{<d>}$, where $< d >$ in the mid-degree of the network. The same steps are iterated over the t number of times or until the required number of opinion leaders identified in the network. Thus, the entire structure of the ROLI algorithm is as follow:

Algorithm 6.2: Reputation-based Opinion Leader Identification (ROLI) Algorithm

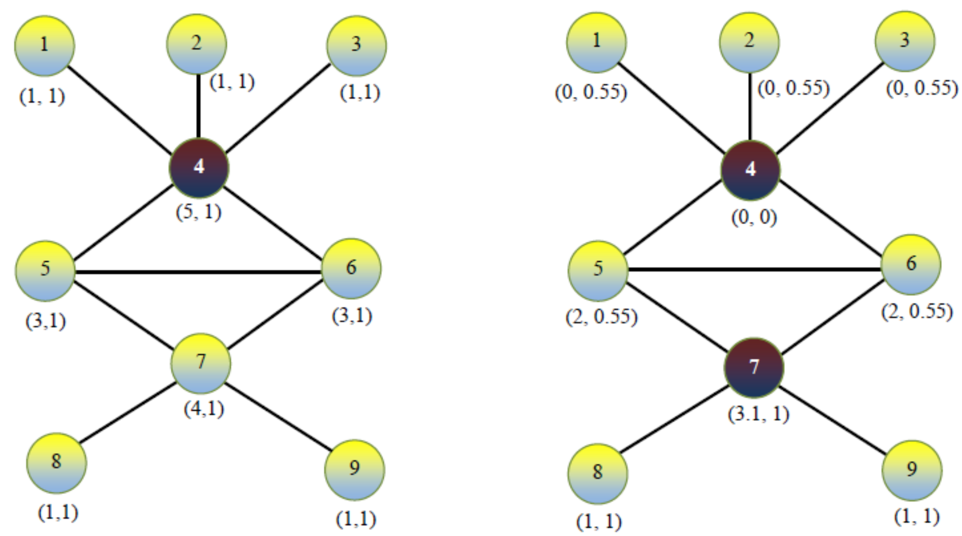
Input: Load the total n number of users

Output: Top-T opinion leaders

Steps:

1. Identified the initial reputation r_x of each node
2. Assign the value to $(vs_x, v_x) \leftarrow (r_x, 1)$ to each node x
3. O[] ← Set of opinion leaders
4. while(i ≤ T)
5.   for x in A do
6.   if x ϵ O:
7.   set $vs_x \leftarrow 0$ and $v_x \leftarrow 0$
   else if (x ϵ N(j) and j ← O)
8.   set $vs_x \leftarrow \sum_{k=1}^{N(x)} r_k$ and $v_x \leftarrow (v_x - f)$
   else:
9.   set $vs_x \leftarrow r_x$ and $v_x \leftarrow 1$
   end if;
10.   find the node j with $\max(vs_j)$ and set O ← O ∪ {j} and A ← A - {j}
11.   i ← i+1
   end for;
 end while;
12. Find the list of top-T opinion leaders return by O
13. end;

To understand the ROLI algorithm more clearly, considered the following network having a total of nine nodes. Initially, each node's voting capacity is set to be one, and the reputation of each node is calculated based on the aggregated degree. .i.e., the total no of links connected with nodes. Fig. 6.6(a) depicts the first round outcomes, and node #4 is elected as an opinion leader. In the next round, all the nodes, i.e., 1,2,3,5, and 6, that are connected with node #4 update their voting ability and voting score as per the mentioned rule. The average degree of the network is 2.22, so the value of f becomes 0.45. therefore, each connected node reduced its voting ability by the factor of f, and the new voting capability of the node becomes 0.55.



(a) (b)

Fig. 6.6. ROLI Algorithm illustration (a) network structure after the first iteration (b) network structure after the second iteration

In the next round, the voting score and voting capability of node #4 become 0, and the reputation of all nodes is measured again. In this phase, node # 7 is selected as the opinion leader with a reputation score of 3.1, as shown in Fig. 6.6(b). the same procedure is carried out for the remaining successive rounds until the designated number of opinion leaders are identified.

6.3 Computational complexity

The process of calculating the Opinion leader identification algorithm's time complexity is divided into three steps; The first step involved the time t1 required to measure the initial reputation and voting ability. The second step included the time t2 needed to select the node with the highest reputation score. The third step had the time t3 required to modify the reputation score. For measuring the initial reputation of each node, total O(n) time is requires. Similarly, the initial voting ability of each node is 1. So, only O(e) time is needed to assign initial voting ability. Thus, the whole time complexity t1 is O(n) + O(e) ≈ O(e). If the optimal procedure is chosen to find the node with the maximum reputation score, the second step's time complexity t2 would be O(n). in the third step, the reputation score of only those nodes would be updated that are one or two units far away from the represser nodes that have been selected as the opinion leaders in the previous round. The mid-degree of the network is $< d >$. Thus the time complexity of step 3 is $O(<d>^2) \approx O\left(\frac{e^2}{n}\right)$. If t number of opinion leaders are selected, the overall complexity would be $O(e+t*n+t\frac{e^2}{n^2})$. If the network is sparse, i.e., n>>e and n>>t, the time complexity of the algorithm might be O(n) for the network.

6.4 Result analysis and performance evaluation

6.4.1 Datasets

6.4.1.1.Twitter dataset

Since the COVID-19 pandemic outbreak, social media have practiced a large volume of data and comments. From the last few time, Twitter also plays an essential role in extracting information about the pandemic. This research has pulled many tweets from Twitter to find out the people’s opinions and sentiments. The main reason behind choosing Twitter is its popularity and usability that has been drastically increased during the pandemic. Some unauthorized sources have posted different kinds of rumors and misinformation. Later on, Twitter has added the fact-checking tie with the tweets to check the tweet's authenticity. So, in this research, we have also extracted Twitter's tweets dataset for analysis purposes. We have used Twitter streaming API to pull out the COVID-19 tweets. Twitter's API supports access to tweets, users, messages, trends, links, and many more things. Due to memory and CPU constraints, we have collected a total of 65.3 M tweets that started from January 1, 2020, to March 30, 2020. We have used the five search keywords for obtaining the tweets. We have gathered the tweet in <Tweet, userId, time, country> format to demonstrate the single tuple.

6.4.1.2 Instagram Dataset

Instagram is another widely used social media used by people for images, videos, and views sharing purposes. In this research, we also have extracted the Instagram posts related to COVID-19. An Instagram API is used for collecting the entire posts started from January 1, 2020, to March 30, 2020. Initially, a token is generated, and POST and GET requests are built through HTTP. However, the total number of posts is relatively more minor than the Twitter data set but contains fascinating facts, rumors, and misinformation about COVID-19. We have collected around 9.8 K posts, and 37 K comments originated from the different validated user accounts.

6.4.1.3 Reddit Dataset

Reddit is an online discussion medium that integrates various things like news, posts, comments, conversations, views, images, and queries. Although the recognition of Reddit is not so wider, yet it covers more rich content and information. Reddit contains lots of communities that provide the information in a very creative and innovative way. We have used the python-based Reddit API to collect the content based on five searching words. We have assembled around 15.7 K comments derived from the various user communities and specific posts. Thus, the statistical description of the datasets is shown in Table 6.1.

Table 6.1. Twitter, Instagram, and Reddit dataset statistical description

Dataset (January 1- March 30, 2020)	**Statistics**		
	Twitter	**Instagram**	**Reddit**
Total number of tweets	65.3 M	46.8 K	25.7 K
% of the tweets in English	71.2%	89.3 %	96.4 %
% of tweets in other and regional languages	28.8%	10.7 %	5.6%
% of verified accounts	8.4%	23.5 %	64.8%
Total number of participating countries	173	148	94
Total number of searched keyword	5 ('Covid19', 'coronavirus', '#2019-ncov', '#covid_19', '#pandemic')		
Density	0.00845	0.00591	0.0139
Clustering coefficient	0.000614	0.000472	0.000863

6.4.2 Analysis and visualization of the experimental result

We have used Gephi, a Java-based network analyzer tool that is used to analyze the network. The tool supports finding out the relation among the users and measuring the network's clustering coefficient. Next, we have used the VADER python library to get the sentiments of each tweet. We have classified each tweet into three classes: positive, negative, and neutral. Further, the reputation of each user is measured based on the aggregate polarity. Trust is also calculated to predict whether the users retweet, forward, or reject the message.

Next, we applied the ROLI algorithm to find the top-T OLs in the network. We have found the list of those user's IDs who have posted maximum tweets or posts about the COVID-19, and other people also commented on those tweets in the network. These tweets contain valuable information and provide a powerful direction towards the controlling from the COVID-19. These tweets are also liked, retweeted, and shared by thousands of users. Further, we have calculated the entropy of each tweet based on measured trust. Here, we have used filtering operation and discard those tweets whose user's reputation score is relatively low. So this operation reduced the whole entropy processing time. Finally, we have obtained the record of users whose tweets are mostly retweeted using SNA measures and suggested reputation score along with their SNA measures as shown in Table 6.2, Table 6.3, and Table 6.4 for all three datasets.

Table 6.2. List of top-10 opinion leaders along with their reputation score and other SNS measures for the Twitter dataset

Node id	DC	Node id	CC	Node id	BC	Node id	PR	Node id	EC	Node id	Reputation
3782784	0.0353824	4882929	0.0099734	5781775	0.0504885	4938103	0.0038952	937482	0.0144372	**758380**	**0.1276728**
1636273	0.0353823	837321	0.0099732	184773	0.0504883	5062765	0.0038951	4287292	0.0144372	**4791216**	**0.1276728**
466738	0.0353821	3877392	0.0099732	2390202	0.0504882	3773291	0.0038949	837174	0.0144371	**829252**	**0.1276727**
1046730	0.0353819	174992	0.0099731	734218	0.0504882	638383	0.0038948	84983	0.0144371	**940251**	**0.1276726**
473721	0.0353815	1062525	0.0099730	78022	0.0504881	194482	0.0038947	2839290	0.014437	**3936037**	**0.1276726**
5254646	0.0353811	494775	0.0099730	519287	0.0504880	2784929	0.0038946	3921043	0.0144369	**7385**	**0.1276726**
904537	0.0353809	105829	0.0099729	3992801	0.0504879	574922	0.0038944	735622	0.0144368	**84892**	**0.1276725**
3029229	0.0353808	2593920	0.0099728	1820378	0.0504877	1383092	0.0038942	1588391	0.0144367	**5827403**	**0.1276725**
2372722	0.0353804	59201	0.0099728	3629048	0.0504876	292739	0.0038941	449293	0.0144367	**1289504**	**0.1276724**
375981	0.0353804	429322	0.0099728	947324	0.0504876	5417321	0.0038941	814871	0.0144366	**683692**	**0.1276724**

Table 6.3. List of top-10 opinion leaders along with their reputation score and other SNS measures for the Instagram dataset

Node id	DC	Node id	CC	Node id	BC	Node id	PR	Node id	EC	Node id	Reputation
3056	0.0954829	8402	0.0703421	17392	0.0639235	19048	0.0418937	8592	0.0390783	**4011**	**0.0838588**
10154	0.0954829	738	0.0703421	9387	0.0639235	7283	0.0418937	5308	0.0390783	**21802**	**0.0838588**

32537	0.0954828	21481	0.0703421	23817	0.0639234	33891	0.0418937	23973	0.0390783	**9820**	**0.0838588**
2098	0.0954828	31412	0.0703420	5927	0.0639234	491	0.0418936	12094	0.0390783	**17391**	**0.0838588**
812	0.0954828	1184	0.0703420	10042	0.0639234	9382	0.0418936	5909	0.0390782	**11896**	**0.0838587**
22904	0.0954827	9823	0.0703420	22893	0.0639234	17495	0.0418936	32984	0.0390782	**36003**	**0.0838587**
4929	0.0954827	37192	0.0703420	387	0.0639233	30874	0.0418936	973	0.0390782	**6298**	**0.0838587**
1090	0.0954826	26177	0.0703419	31903	0.0639233	5983	0.0418935	128	0.0390782	**25009**	**0.0838586**
25709	0.0954826	2017	0.0703419	2851	0.0639233	24981	0.0418935	31983	0.0390781	**14892**	**0.0838586**
8341	0.0954826	18451	0.0703419	7389	0.0639232	13722	0.0418935	5582	0.0390781	**439**	**0.0838586**

Table 6.4. List of top-10 opinion leaders along with their reputation score and other SNS measures for the Reddit dataset

Node id	**DC**	**Node id**	**CC**	**Node id**	**BC**	**Node id**	**PR**	**Node id**	**EC**	**Node id**	**Reputation**
1906	0.1073821	369	0.0852855	4329	0.1019903	6525	0.0753870	9241	0.0726944	**8033**	**0.1080275**
5648	0.1073821	6407	0.0852855	16132	0.1019903	12599	0.0753870	5481	0.0726944	**7075**	**0.1080275**
21653	0.1073821	918	0.0852855	2621	0.1019903	3745	0.0753870	5575	0.0726944	**11575**	**0.1080275**
1842	0.1073820	8212	0.0852855	11921	0.1019902	14492	0.0753870	942	0.0726943	**18284**	**0.1080275**
13654	0.1073820	13734	0.0852854	20729	0.1019902	5999	0.0753869	8219	0.0726943	**13057**	**0.1080274**
5608	0.1073820	6566	0.0852854	1022	0.1019902	20562	0.0753869	1926	0.0726943	**5402**	**0.1080274**
4251	0.1073820	6962	0.0852854	3799	0.1019902	2224	0.0753869	21107	0.0726942	**1099**	**0.1080274**
10648	0.1073819	19048	0.0852854	407	0.1019901	8974	0.0753869	1601	0.0726942	**12056**	**0.1080274**
1149	0.1073819	1529	0.0852854	11089	0.1019901	10425	0.0753869	180	0.0726942	**5148**	**0.1080273**
8457	0.1073819	8979	0.0852853	931	0.1019901	19616	0.0753868	12621	0.0726942	**9236**	**0.1080273**

Once we identified the top-T(=10) OLs in the network, the next step is to choose the threshold value for the entropy. It is critical to select a particular threshold value λ for declaring a tweet as a rumor. Therefore, after various analyses and experiments, we have chosen λ=0.95, i.e., if the entropy of the tweet is less than 0.95, the tweet would be reported as a rumor; otherwise, the OL may only forward or add their comments with the post and forward to the followers and other users in the network.

6.5 Performance metrics for rumor controlling

To evaluate the performance and effectiveness of the proposed approach for COVID-19 rumor controlling, we have used the performance metrics based on the behavior of the oblivious node n_o, diffuser node n_d, and represser node n_r. As we have mentioned previously, the rumor spreading tendency is similar to the disease spreading behavior in the traditional SIR epidemic model. The authenticity of the proposed approach depends on how quickly it identifies and controls the total number of rumors in the network. Thus, we have used three metrics: diffuser degree, represser degree, and affected degree, respectively, for measuring the approach's performance.

6.5.1 Diffuser degree

The diffuser degree explains the behavior of the diffuser node in the network. If the rumors spreading rate β is high, most of the nodes might be infected or influenced by the rumors. It is mandatory to transmit the correct information in the network as early as possible to avoid any crisis. So, The diffuser degree $f_d(t)$ is defined as the ratio between the total number of diffuser nodes and the total count of represser nodes, oblivious nodes, and diffuser nodes in the network at time *t* as shown in Eq.(6.14).

$$f_d(t) = \frac{n_d}{n_d+n_r+n_o} \quad \dots (6.14)$$

In Fig. 6.7, we can observe that as soon as the count of OLs increased, the diffuser degree reduced gradually; But due to some spreader nodes with a strong belief in the rumor, it is impossible to reach the zero level. Also, the proposed approach reduced the total number of diffusers 26% faster than other SNA measures as the number of OLs increased gradually.

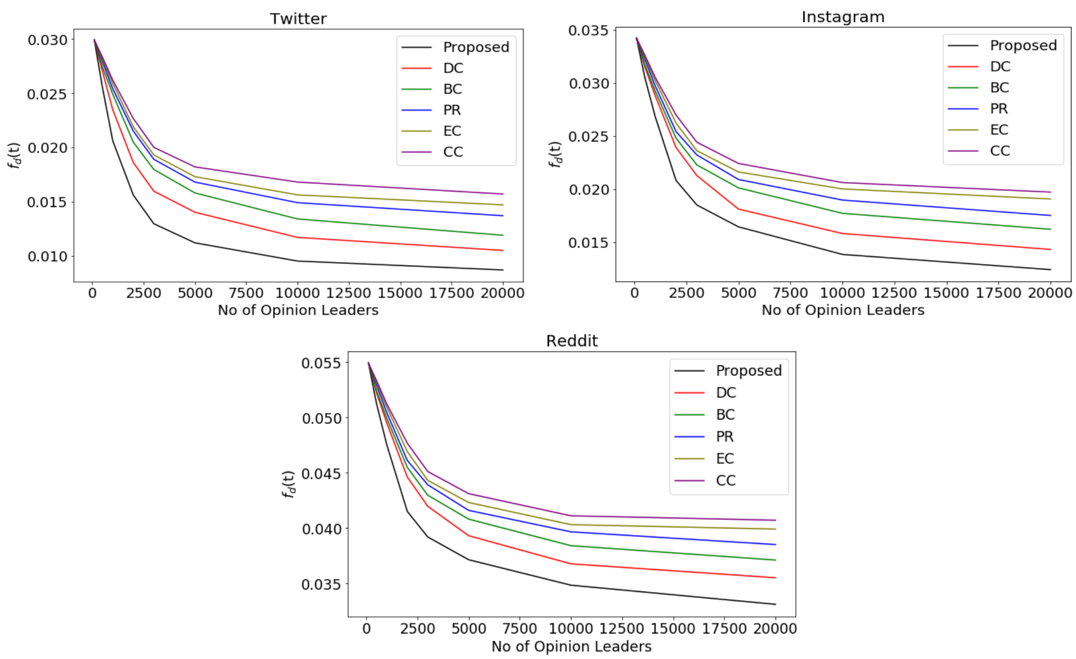


Fig. 6.7. Visualization of diffuser degree for proposed approach and standard SNA measures for Twitter, Instagram, and Reddit

6.5.2 Represser degree

Represser degree states the control of the represser node in the network because eventually, the represser nodes are considered the OLs in the network. So, if the represser nodes identify the tweet's trueness as soon as possible, they can help stop the rumor from spreading. Thus, represser degree $f_r(t)$ is defined as the ratio between the total number of represser nodes and the total count of represser nodes, oblivious nodes, and diffuser nodes in the network at time *t* as shown in Eq.(6.15).

$$f_r(t) = \frac{n_r}{n_d + n_r + n_o} \quad \dots (6.15)$$

In Fig. 6.8, we can monitor the relationship between the represser degree and diffuser degree over time *t*. As soon as the represser nodes spread the actuality of the rumor with the rate *γ*, the represser degree increased, and the depressor degree decreased with time. Again, the proposed approach performed better and spreading veracity around 22% faster than other SNA measures.

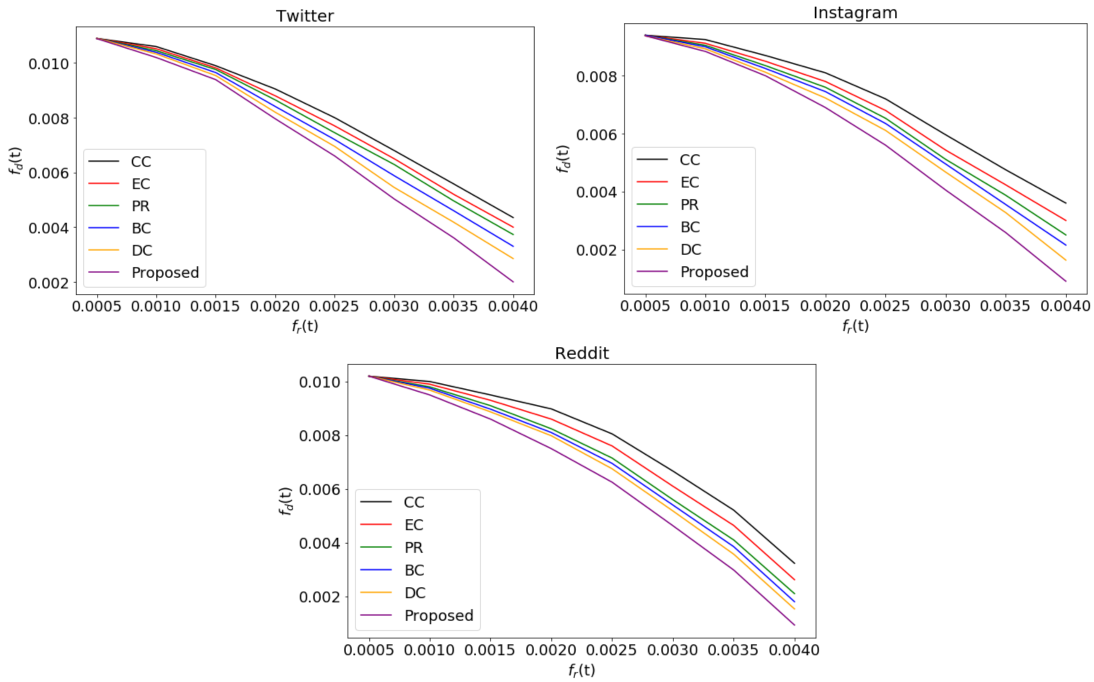


Fig. 6.8. Visualization of represser degree for proposed approach and standard SNA measures for Twitter, Instagram, and Reddit

6.5.3 Affected degree

An affected degree is a scale that explains the effect of OLs in the network. It measures the total number of users who are influenced by OLs. The represser degree $f_a(t)$ is defined as the fraction between the total number of influenced nodes and the total count of represser nodes, oblivious nodes, diffuser nodes, and affected nodes in the network at time *t*, as shown in Eq.(6.16).

$$f_a(t) = \frac{n_a}{n_d+n_r+n_o+n_a} \quad \dots (6.16)$$

Where n_a depicts the number of users influenced by the OLs. Further in Fig. 6.9, we can infer that as represser nodes spread the rumor's actuality with the rate *γ*, the affected degree increases with time, i.e., more number of users influenced by the OLs. The proposed approach produced better outcomes and impacted the users approximately 23% faster than other SNA measures.

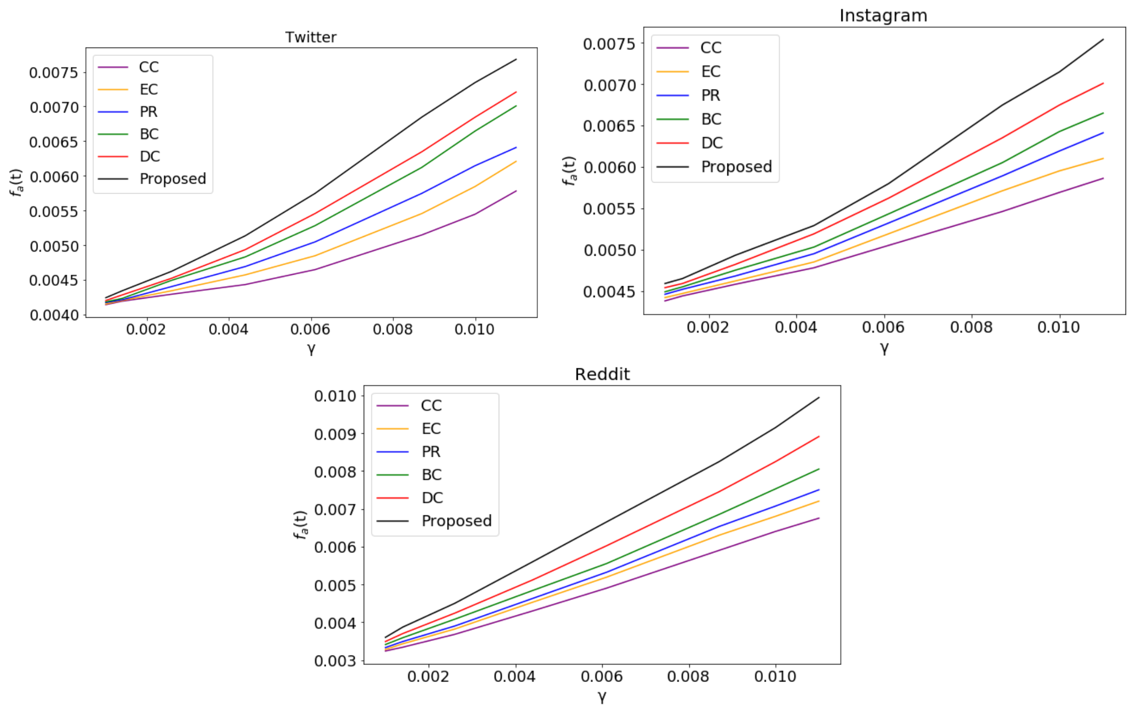
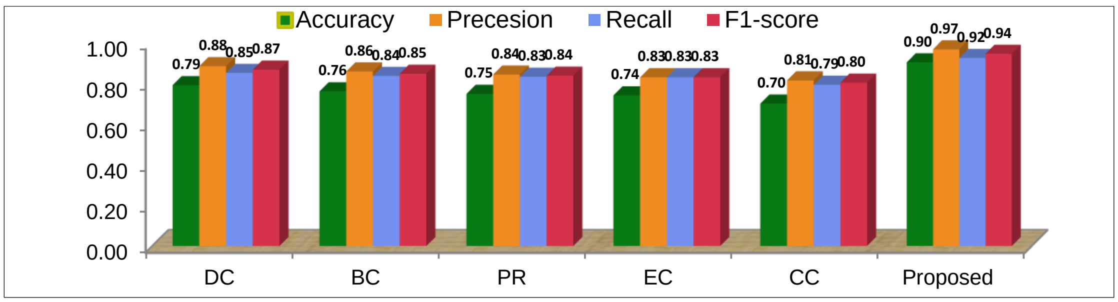


Fig. 6.9. Visualization of affected degree for proposed approach and standard SNA measures for Twitter, Instagram, and Reddit

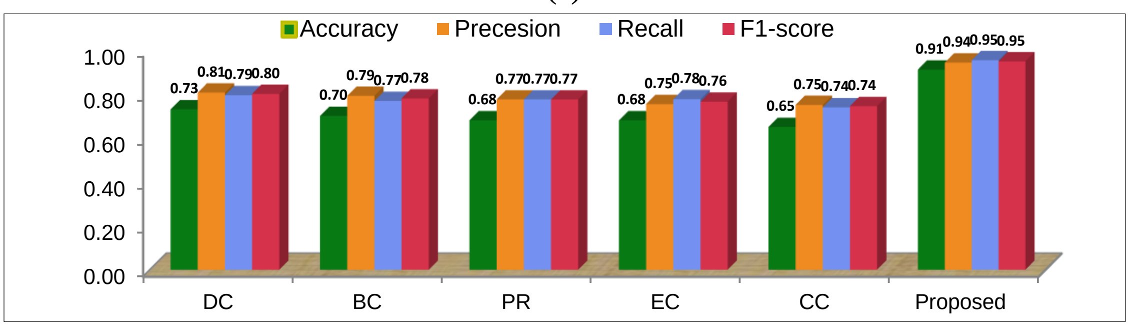
OLs are identified at different moments according to the approach, i.e., as soon as the time elapsed, the number of OLs also increased as well. Each OL has a different number of followers, and rumor spreaders are also presented in the network. So, the intensity to control the rumors among the general users varies and depends on the total number of followers and network structure.

6.5.4 Performance metrics for ROLI algorithm

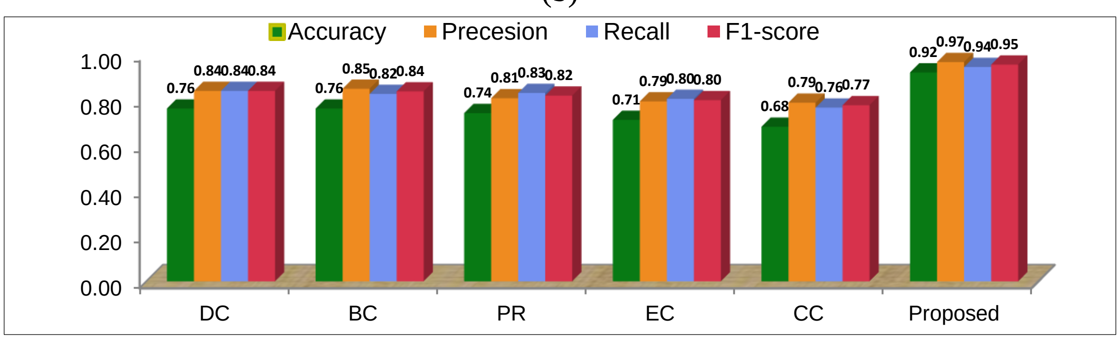
In this research, we have proposed a novel ROLI algorithm for OLs recognition in OSN. To validate the ROLI algorithm's performance, we have compared the results with the other SNA measures. We compared the impact of the proposed ROLI algorithm with the standard SNA measures, which are also used globally to find the prominent users in the network. we utilized the four performance metrics; Accuracy, Precision, Recall, and F1-score, to ensure the reaching of the approach [234]. true-positive(TP), true-negative (TN), false-positive(FP), and false-negative(FN) components help to compute these metrics. In Fig. 6.10, we have compared the outcomes turn out from the proposed ROLI algorithm w.r.t. mentioned performance metrics. We have observed that the proposed algorithm gave around 91% accuracy,93% precision, 95% recall, and 94% F1-score. For experiment purpose, we have set the value of $\beta = 0.0055$ and $\gamma = 0.0085$. So, all the results and experiments have been performed based on these parameters.



(a)



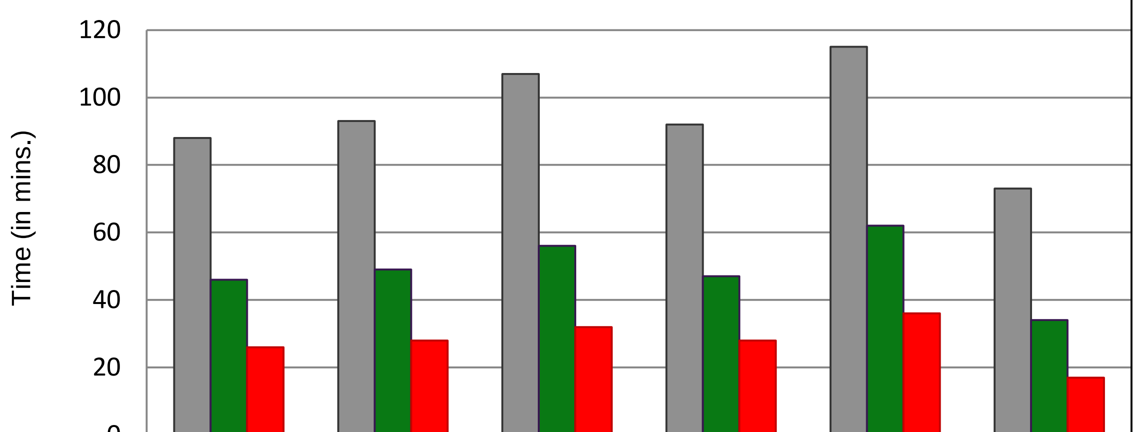
(b)



(c)

Fig. 6.10. Analysis of Performance metrics between Proposed ROLI and standard SNA measures for (a) Twitter, (b) Instagram, and (c) Reddit dataset

Another significant aspect that also affects the importance of the algorithm is the execution time. In the previous section, we discussed the algorithm's complexity, which is almost equivalent to O(n). We have considered only that time that is consumed to find out the designated OLs from the network. Each method has its parameters to estimate the OLs are very materialistic. In Fig. 6.11, we have demonstrated the execution time required by each SNA measure along with the proposed ROLI algorithm. From the analysis, we examined that the proposed algorithm needed a shorter time that is comparatively lower than the other SNA measures for all three datasets. In our approach, we have also performed extra work to compute the reputation, trust, and entropy, which also needed additional time. Thus, we can strongly recommend that the proposed algorithm perfectly controls COVID-19 rumors and supports strong beliefs and reputation on the social network.



	DC	BC	PR	EC	CC	Proposed
Twitter	88	93	107	92	115	73
Instagram	46	49	56	47	62	34
Reddit	26	28	32	28	36	17

Fig. 6.11. Execution time analysis of ROLI algorithm with standard SNA measures

6.6 Chapter summary

In the present pandemic condition, the COVID-19 disease has been ultimately affected the whole world drastically. The role and power of social media are very vital regarding COVID-19 related rumors and misinformation propagation. Although the World Health Organization(WHO) and other official government organizations already circulated various guidelines and control measures to avoid the disease, different kinds of misinformation being spread by numerous sources in social media [235], [236]. Thus, it is essential to control such rumors and misinformation to save public health. Since Twitter's popularity is drastically increased after this pandemic, consumers have posted different kinds of information without checking the source authenticity.

In this research, an innovative approach is addressed that can control the COVID-19 related rumors on social media up to certain limits as much as possible. First, many tweets are extracted from the Twitter, Instagram, and Reddit social networks for analysis purposes. After prepossessing the tweets, Real-time monitoring of tweets, rumors, patterns, trends, and misinformation is required to avoid any trouble. Such control may identify only the verified information and ensure that only verified and trusted information would be transmitted to control dangerous consequences.

Chapter 7

Conclusion and future scope

This chapter coats the complete summary of the proposed various approaches related to opinion leaders' identifications, power, and applications in online social networks. Sections 8.1 cover the outline of all the methods; section 8.2 covers the limitations of the proposed methods. Section 8.3 explores the future aspects of the proposed work, and finally, the concluding interpretations of the research are wrapped up in section 8.4.

7.1 Research summary

This study about the opinion leaders presented unique and novel approaches to detect the opinion leader, explore their power for information diffusion, and demonstrate opinion leaders' applicability in the healthcare domain to prevent COVID-19 rumors in the online social networks. As per the literature, various studies found opinion leaders but have some constraints and limited accuracy. The social network's biggest challenge is the dataset's size and the unavailability of a specific and accurate dataset. Although various SNA measures are used for comparative analysis and research purposes, no other benchmark methods are present to find the opinion leaders in the online social network.

For fulfilling the need of RO1, two nature-inspired metaheuristic algorithms discover the opinion leaders. For many years, meta-heuristic algorithms can solve various real-world problems efficiently and effectively. The first approach is based on the nature-inspired Firefly algorithm. The concept of the algorithm is based on the global community and blinking actions of Firefly. The second approach is based on the Whale optimization problem. The meta-heuristic nature-inspired social network-based Whale Optimization Algorithm (SNWOA) is proposed. Each user behaves as the whale that tries to find out the other user having more reputation. The user has more importance considered prey, and all the users wanted to connect to the user with more prominence. Both approaches are incredibly efficient in discovering the opinion leader with high accuracy. In some cases, the performance of whale optimization is better than the firefly algorithm. Chapter 2 covers the whole description of both the algorithms in detail.

For achieving the RO2, two solutions have been proposed. In the first solution, a game theory-based approach is addressed that hypothesized each user like a player who wants to make coalitions with the other users in the network. Trust and centralities considered as attributes help to find the marginal contribution of a user in the game. Shapely value is also used to calculate the marginal contribution. Four different solutions are proposed to compute individual payoff. The proposed approach is convenient for information diffusion because companies collaborate for strategic growth and benefits. Another solution is based on Graph Neural Network that consists of the GOLI model utilizing the power of GNN to categorize the opinion leaders. The main target is to develop an approach that captures the hidden and latent information from the network. The accuracy of the deep learning-based proposed model is very high and can effectively support information diffusion.

To attain the RO3, a new approach is delivered consisting of opinion leaders' applicability and power to stop COVID-related rumors in the online social network. In leading motive is to find out the relevance of opinion leaders through online social networks in healthcare. In the ongoing COVID-19 pandemic, people spread various COVID-19 related rumors and hoaxes, which incredibly negatively influences civilization. Due to these rumors, there are lots of mental illnesses, and anxieties arise. Thus, a rumor prevention approach is addressed in which opinion leaders are selected to validate the tweet's integrity based on entropy measurement and sentiment analysis. The proposed approach's performance is compared based on three metrics; diffuser degree, represser degree, and affected degree, respectively.

7.2 Theoretical and practical implications of research

7.2.1 Theoretical implications

Nowadays, most of the human decision-making process is influenced by social media activities. This research produces innovative and novel implications representing the identification of opinion leaders and the impact on information diffusion and controlling rumors. In the entire study, various statistical, mathematical, and scientific calculations are used, which support all the analysts and mathematicians to understand the power of graph theory.

In the research, novel nature-inspired meta-heuristics algorithms, game theoric approaches, and centrality-based methods are used to identifying the opinion leader in the online social network. So this research is very significant for researchers interested in taking insight knowledge of nature-inspired algorithms and opening the new direction towards opinion leader identification.

Opinion leaders' importance is vital and impactful for managing various rumors and misinformation transmitted over online social networks. In this research, we have explored the power of opinion leaders to control COVID-19 rumors. Thus, this research's contribution is very significant in various domains to prevent rumors and hoaxes comprehensively. In previous studies, only limited approaches used reputation and trust to find opinion leaders in limited environments. The main essence of this study is the utilization of reputation and trust effectively. Also, the polarity of each tweet is measured to identify the sentiments of the population about the COVID-19 epidemic.

In this study, we did not apply any data mining approach. Although data mining approaches outperformed well in real-world applications and produced improved outcomes, they are not appropriate for online social networks due to their dynamic nature. In this study, we have merely used software tools, statistical formulas, and methods to discover opinion leaders and validate the integrity of the post. One of the principal merits of this approach is that it is very beneficial for large datasets. As the number of users steadily increased in the network, more users would exchange their opinions and views. So, at that moment, each user can measure other users' reputations, attractiveness, and degree of trust more accurately and precisely.

The proposed system's complexity is straightforward and does not include any multifaceted structure and formulation. For all operations, elementary strategies are used. Most of the previously developed methods involved complex, composite, and lengthy computations that are difficult to implement and understand.

This study makes exceptional progress towards innovation in information science by exploring and demonstrating the power of opinion leaders to information diffusion and control COVID-19 rumors in the current pandemic. Also, the operational behavior of the proposed ROLI approach is analogous to the SIR epidemic model that shows the spreading of disease in the real world. Such types of research also support preventing a high amount of misinformation transition in online social networks. It is essential to identify the set of users who have an immense impact on their followers to improve the trustworthiness of online social networks. So this research contributes enormously to fulfill this objective with higher accuracy and effectiveness.

7.2.2 Practical implications

In this research, we have identified the top-N opinion leaders who can diffuse the information and control the transmission of rumors as much as possible. This research's practical implication is very much in the real world because the diffusion of any negative news may develop the mystification related to COVID-19 treatment and diagnosis that affect human health and inner sentiment support system [237]. The influence of opinion leaders is limited to preventing misinformation or rumors and is very significant in various domains like education, agriculture, healthcare, defense, marketing, promotions, consumer behavior, and more. The practical implication of this research is broadly elaborated as follows.

The research outcomes are highly improved w.r.t. accuracy, precision, recall, F1-score, and needed lesser execution time than other standard SNA measures for all the datasets. Previous studies also computed the opinion leaders based on different parameters. Still, only limited studies used the integration of reputation, centrality, trust, polarity, and entropy to determine opinion leaders in online social networks. The impact of the proposed techniques is considerably picked up over other SNA measures, as presented in the previous section.

Different types of datasets are chosen for analysis and evaluation purposes in this research because of their nature and dynamics. Twitter has a retweet facility that is not presented on Instagram. Instagram is an attention-based social network without any messy content. Reddit is a forum-based social network where a user can taste the essence of all kinds of information. The density of Reddit is higher than both of the networks but having a limited number of users. So, this work outperformed well for all kinds of online social networks. Previous studies used only datasets suitable to a specific domain or interest and did not cover various online social networks.

Third, there are various real-world insinuations of this research. It is a unique approach supporting organizations and industries to identify their customer's sentiments and opinions about a specific product through online social networks. In healthcare, opinion leaders can find healthcare experts, physicians, and advisers through online physician communities. In agriculture, opinion leaders can support farmers by providing advanced practices and propagating different pesticide information, leading to an enhanced sustainable agricultural system. Opinion leaders also help organizations differentiate between rumors and anti-rumor messages on online social network through the proposed approach by calculating the message's entropy. Similarly, the proposed system performed well for opinion leader detection and control rumors in multiple fields and provinces.

7.3 Limitations of the work

No one is perfect in the world, and often, each research has some limitations and barriers. This work also has the following restrictions:

- One of the biggest challenges with the social network is the dataset's size because of the infinite amount of data recurrently posted by millions of users daily. In this research, only static data set is used for analysis purposes due to the unavailability of the dynamic data set [96], [238].

- The other limitation is that only elementary SNA measures are used for comparative analysis and evaluation. Some other SNA measures can also be used for analysis purposes due to the technological revolution and inventions.
- Randon selection algorithm is used for choosing the opinion leaders as seed users for information diffusion.
- Only limited vital words are chosen to extract the tweets due to space constraints for COVID-19 rumor prevention.

7.4 Future aspects

Following are the future perspective of the work.

- Other nature-inspired metaheuristic techniques and measuring tools can be discovered to visualize and interpret opinion leaders' recognition techniques by analyzing a real-time dynamic dataset [160], [239] .
- Some more advanced deep-learning-based models are associated with other user's multi-relational characteristics such as user's response time, geographical location, the domain of interest, etc., as upcoming directives for computing [240], [241].
- The innovative computational intelligence techniques along with the evolutionary game theory approach can be used to detect the promising opinion leaders in online social networks [70].
- The applicability of opinion leaders might be explored in agriculture, defense, disaster management, medical, recommender system, and other critical domains [242]–[246].
- Some other variations of the SNA measure can be used for comparative analysis [247].
- A model can be designed that can supervise the tweets and their origin dynamically under the supervision of opinion leaders so that other people would be more aware of the COVID-19 rumors.

- The power of opinion leaders might be adequate to resolve the cold start problem in the recommender system [248].
- The work can be extended to find opinion leaders using user's posts like images, audios, videos, etc.
- Other social networks, like Facebook, YouTube, Flicker, Instagram, etc. can also be utilized for extracting the dataset to find the opinion leaders [249], [250].

7.5 Chapter summary

The applicability of the proposed methods is very extensive and useful. Nowadays, every organization wants to create wealth in the commercial market. Each company desires that the sale of its products would be high in the future. So opinion leaders have significant exposure to achieve this goal. In the global industry and the other fields as education, healthcare, agriculture, e-commerce, digital marketing, population controlling, and many more. Recently, lots of research has been developed to illustrate opinion leaders' felicitousness to solve many real-world problems. The pertinence of all the techniques would also be appreciably helpful in the diverse other modern applications such as catalyst analysis in the chemical reaction, GIS, biological pattern, modern algebra, regression analysis, cloud computing, load balancing, distributed computing, and many more as an influencer, key leader, opinion booster, opinion miner, key miner, candidate user, change agent, and other deviated forms [251].

References

[1] J. A. Barnes, "Class and Committees in a Norwegian Island Parish," *Hum. Relations*, vol. 7, no. 1, 1954, doi: 10.1177/001872675400700102.

[2] S. Milgram, "small world problem," *Psychol. Today*, vol. 1, no. 1, 1967.

[3] B. Howard, "Analyzing Online Social Networks," *Commun. ACM*, vol. 51, no. 11, pp. 14–16, Nov. 2008, doi: 10.1145/1400214.1400220.

[4] J. Travers and S. Milgram, "An experimental study of the small world problem," in *The Structure and Dynamics of Networks*, vol. 9781400841356, 2011.

[5] A. Degenne and M. Forse, *Introducing Social Networks*. 2011.

[6] E. Ferrara, "Mining and Analyzing Online Social Networks," *Soc. Networks*, no. April, 2011.

[7] M. Parameswaran and A. B. Whinston, "Research issues in social computing," *J. Assoc. Inf. Syst.*, vol. 8, no. 6, 2007, doi: 10.17705/1jais.00132.

[8] E. Lazega, "The collegial phenomenon: a structural theory of collective action among peers.," *Rev. Fr. Sociol.*, vol. 40, no. 4, 1999.

[9] F. Y. Wang, D. Zeng, K. M. Carley, and W. Mao, "Social computing: From social informatics to social intelligence," *IEEE Intelligent Systems*, vol. 22, no. 2. 2007, doi: 10.1109/MIS.2007.41.

[10] R. Miliband, K. Marx, T. B. Bottomore, and M. Rubel, "Karl Marx. Selected Writings in Sociology and Social Philosophy," *Br. J. Sociol.*, vol. 8, no. 4, 1957, doi: 10.2307/587984.

[11] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, "Microscopic evolution of social networks," 2008, doi: 10.1145/1401890.1401948.

[12] B. Krishnamurthy and C. E. Wills, "On the leakage of personally identifiable information via online social networks," 2009, doi: 10.1145/1592665.1592668.

[13] E. Buechel, "Online Social Networks : Why Do We Use Online Social Networks and How Do They Affect Us ?," *Adv. Consum. Res.*, vol. 40, 2012.

[14] P. Y. Chiu, C. M. K. Cheung, and M. K. O. Lee, "Online social networks: Why do 'we' use facebook?," in *Communications in Computer and Information Science*, 2008, vol. 19, doi: 10.1007/978-3-540-87783-7_9.

[15] S. Wasserman and K. Faust, "Social Network Analysis: Methods and Applications," *Struct. Anal. Soc. Sci. vol.8*, 1994, doi: 10.1525/ae.1997.24.1.219.

[16] R. Klamma and C. Haasler, "Wikis as Social Networks: Evolution and Dynamics," *2nd SNA-KDD Work. Soc. Netw. Min. Anal.*, 2008.

[17] G. Kukła, P. Kazienko, P. Bródka, and T. Filipowski, "Recommendation boosted query propagation in the social network," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2010, vol. 6430 LNCS, doi: 10.1007/978-3-642-16567-2_9.

[18] W. S. Yang, H. C. Cheng, J. Ben Dia, and H. T. Lin, "Mining social networks for

targeted advertising," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2006, vol. 6, doi: 10.1109/HICSS.2006.272.

[19] C. A. Yeung, I. Liccardi, K. Lu, O. Seneviratne, and T. Berners-lee, "Decentralization : The Future of Online Social Networking," *Artif. Intell.*, vol. 2, 2006.

[20] L. A. Adamic and E. Adar, "Friends and neighbors on the Web," *Soc. Networks*, vol. 25, no. 3, 2003, doi: 10.1016/S0378-8733(03)00009-1.

[21] G. W. Flake, S. Lawrence, and C. L. Giles, "Efficient identification of web communities," 2000, doi: 10.1145/347090.347121.

[22] B. Wellman, J. Salaff, D. Dimitrova, L. Garton, M. Gulia, and C. Haythorntkwaite, "Computer networks as social networks: Collaborative Work, Telework, and Virtual Community," *Annu. Rev. Sociol.*, vol. 22, 1996, doi: 10.1146/annurev.soc.22.1.213.

[23] J. Golbeck and J. Hendler, "FilmTrust: Movie recommendations using trust in Web-based social networks," in *2006 3rd IEEE Consumer Communications and Networking Conference, CCNC 2006*, 2006, vol. 1, doi: 10.1109/CCNC.2006.1593032.

[24] D. M. Boyd and N. B. Ellison, "Social network sites: Definition, history, and scholarship," *J. Comput. Commun.*, vol. 13, no. 1, pp. 210–230, 2007, doi: 10.1111/j.1083-6101.2007.00393.x.

[25] J. A. Obar and S. Wildman, "Social media definition and the governance challenge: An introduction to the special issue," *Telecommunications Policy*, vol. 39, no. 9. 2015, doi: 10.1016/j.telpol.2015.07.014.

[26] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," *Bus. Horiz.*, vol. 53, no. 1, 2010, doi: 10.1016/j.bushor.2009.09.003.

[27] M. Foster, "Why Users Participate in Online Social Networks," *Int. J. E-bus. Manag.*, vol. 4, no. 1, 2010, doi: 10.3316/ijebm0401003.

[28] R. E. Wilson, S. D. Gosling, and L. T. Graham, "A Review of Facebook Research in the Social Sciences," *Perspect. Psychol. Sci.*, vol. 7, no. 3, 2012, doi: 10.1177/1745691612442904.

[29] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," 2007, doi: 10.1145/1298306.1298311.

[30] P. Parau, C. Lemnaru, and R. Potolea, "Assessing vertex relevance based on community detection," in *IC3K 2015 - Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 2015, vol. 1, doi: 10.5220/0005596300460056.

[31] D. Gold, E. Katz, P. F. Lazarsfeld, and E. Roper, "Personal Influence: The Part Played by People in the Flow of Mass Communications.," *Am. Sociol. Rev.*, vol. 23, no. 4, pp. 572–574, 1956, doi: 10.2307/2088435.

[32] S. Y. Song, E. Cho, and Y. K. Kim, "Personality factors and flow affecting opinion leadership in social media," *Pers. Individ. Dif.*, 2017, doi: 10.1016/j.paid.2017.03.058.

[33] P. W. Turnbull and A. Meenaghan, “Diffusion of Innovation and Opinion Leadership,” *Eur. J. Mark.*, vol. 14, no. 1, 1980, doi: 10.1108/EUM0000000004893.

[34] T. W. Valente, “Social network thresholds in the diffusion of innovations,” *Soc. Networks*, 1996, doi: 10.1016/0378-8733(95)00256-1.

[35] S. A. Semenkovich and O. A. Tsukanova, “On the Algorithms of Identifying Opinion Leaders in Social Networks,” 2019, doi: 10.1016/j.procs.2019.12.050.

[36] Y. Zhao, G. Kou, Y. Peng, and Y. Chen, “Understanding influence power of opinion leaders in e-commerce networks: An opinion dynamics theory perspective,” *Inf. Sci. (Ny).*, 2018, doi: 10.1016/j.ins.2017.10.031.

[37] R. Iyengar, C. Van den Bulte, and T. W. Valente, “Opinion Leadership and Social Contagion in New Product Diffusion,” *Mark. Sci.*, vol. 30, no. 2, pp. 195–212, 2011, doi: 10.1287/mksc.1100.0566.

[38] K. K. Chan and S. Misra, “Characteristics of the opinion leader: A new dimension,” *J. Advert.*, 1990, doi: 10.1080/00913367.1990.10673192.

[39] P. Parau, C. Lemnaru, M. Dinsoreanu, and R. Potolea, “Opinion Leader Detection,” in *Sentiment Analysis in Social Networks*, 2016.

[40] V. P. RICHMOND, “MONOMORPHIC AND POLYMORPHIC OPINION LEADERSHIP WITHIN A RELATIVELY CLOSED COMMUNICATION SYSTEM,” *Hum. Commun. Res.*, vol. 6, no. 2, 1980, doi: 10.1111/j.1468-2958.1980.tb00131.x.

[41] G. Doumit, F. C. Wright, I. D. Graham, A. Smith, and J. Grimshaw, “Opinion leaders and changes over time: A survey,” *Implement. Sci.*, vol. 6, no. 1, 2011, doi: 10.1186/1748-5908-6-117.

[42] T. Rapanos, “What Makes an Opinion Leader: Expertise Versus Popularity,” *SSRN Electron. J.*, 2019, doi: 10.2139/ssrn.3454606.

[43] T. W. Valente and P. Pumpuang, “Identifying opinion leaders to promote behavior change,” *Heal. Educ. Behav.*, 2007, doi: 10.1177/1090198106297855.

[44] E. Kuznetsova and E. Zinovyeva, “Psychological aspects of identifying and studying digital opinion leaders,” *World Sci. Pedagog. Psychol.*, vol. 8, no. 5, 2020, doi: 10.15862/97psmn520.

[45] J. Turcotte, C. York, J. Irving, R. M. Scholl, and R. J. Pingree, “News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking,” *J. Comput. Commun.*, 2015, doi: 10.1111/jcc4.12127.

[46] F. Meng and J. Wei, “Research on the influential factor of consumer model based on online opinion leader,” *Comput. Model. New Technol.*, vol. 18, no. 7, 2014.

[47] Y. Li, S. Ma, Y. Zhang, R. Huang, and Kinshuk, “An improved mix framework for opinion leader identification in online learning communities,” *Knowledge-Based Syst.*, vol. 43, pp. 43–51, May 2013, doi: 10.1016/j.knosys.2013.01.005.

[48] M. jie Liao, J. Zhang, R. mei Wang, and L. Qi, “Simulation research on online marketing strategies of branded agricultural products based on the difference in opinion leader attitudes,” *Inf. Process. Agric.*, 2020, doi: 10.1016/j.inpa.2020.12.001.

[49] F. Li and T. C. Du, “Listen to me - Evaluating the influence of micro-blogs,” *Decis.*

Support Syst., vol. 62, 2014, doi: 10.1016/j.dss.2014.03.008.
[50] M. Shinde and S. Girase, “Identification of topic-specific Opinion Leader using SPEAR algorithm in Online Knowledge communities,” 2017, doi: 10.1109/CAST.2016.7914956.
[51] L. Zhu, G. Yin, and W. He, “Is this opinion leader’s review useful? Peripheral cues for online review helpfulness,” *J. Electron. Commer. Res.*, vol. 15, no. 4, 2014.
[52] S. A. Mohammadi and A. Andalib, “Using the opinion leaders in social networks to improve the cold start challenge in recommender systems,” in *2017 3rd International Conference on Web Research, ICWR 2017*, 2017, pp. 62–66, doi: 10.1109/ICWR.2017.7959306.
[53] X. Song, Y. Chi, K. Hino, and B. Tseng, “Identifying opinion leaders in the blogosphere,” in *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management - CIKM ’07*, 2007, pp. 971–974, doi: 10.1145/1321440.1321588.
[54] R. Van Der Merwe and G. Van Heerden, “Finding and utilizing opinion leaders : Social networks and the power of relationships,” *South African J. Bus. Manag.*, vol. 40, no. 3, pp. 65–77, 2009, doi: 10.4236/sn.2013.21002.
[55] S. M. H. Bamakan, I. Nurgaliev, and Q. Qu, “Opinion leader detection: A methodological review,” *Expert Systems with Applications*, vol. 115. pp. 200–222, 2019, doi: 10.1016/j.eswa.2018.07.069.
[56] S. Winter and G. Neubaum, “Examining Characteristics of Opinion Leaders in Social Media: A Motivational Approach,” *Soc. Media Soc.*, vol. 2, no. 3, 2016, doi: 10.1177/2056305116665858.
[57] L. G. Corey, “People who Claim to be Opinion Leaders: Identifying their Characteristics by Self-report,” *J. Mark.*, vol. 35, no. 4, 1971, doi: 10.1177/002224297103500409.
[58] T. E. Ness, D. L. Loudon, and A. J. Della Bitta, “Consumer Behavior: Concepts and Applications,” *J. Mark. Res.*, vol. 17, no. 1, 1980, doi: 10.2307/3151130.
[59] F. Li and T. C. Du, “Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs,” *Decis. Support Syst.*, vol. 51, no. 1, pp. 190–197, 2011, doi: 10.1016/j.dss.2010.12.007.
[60] H. C. Lin, P. F. Bruning, and H. Swarna, “Using online opinion leaders to promote the hedonic and utilitarian value of products and services,” *Bus. Horiz.*, 2018, doi: 10.1016/j.bushor.2018.01.010.
[61] M. P. Venkatraman, “Opinion Leadership, Enduring Involvement and Characteristics of Opinion Leaders: A Moderating or Mediating Relationship,” *Adv. Consum. Res.*, vol. 17, no. 1, 1990.
[62] K. Simbeck, “Opinion Leaders and Lead Users in Marketing and Management Accounting and Impact on Small Business Performance,” *Soc. Netw.*, vol. 02, no. 01, 2013, doi: 10.4236/sn.2013.21002.
[63] S. J. Tsang and H. Rojas, “Opinion Leaders, Perceived Media Hostility and Political Participation,” *Commun. Stud.*, 2020, doi: 10.1080/10510974.2020.1791203.
[64] G. Flodgren, M. A. O’brien, E. Parmelli, and J. M. Grimshaw, “Local opinion

leaders: Effects on professional practice and healthcare outcomes," *Cochrane Database of Systematic Reviews*, vol. 2019, no. 6. 2019, doi: 10.1002/14651858.CD000125.pub5.
[65] "Research on the Role of 'Opinion Leaders' in College Students' Ideological and Political Education," 2018, doi: 10.23977/icettl.2018.71129.
[66] T. S. Hameed and B. Sawicka, "The Importance of Opinion Leaders in Agricultural Extension," *World Sci. News*, vol. 76, 2017.
[67] Y.-S. Kwon and H. Ryong Song, "The Role of Opinion Leaders in Influencing Consumer Behaviors with a Focus on Market Mavens: A Meta-analysis," *ATHENS J. MASS MEDIA Commun.*, vol. 1, no. 1, 2014, doi: 10.30958/ajmmc.1-1-3.
[68] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: A survey," *Knowledge-Based Syst.*, vol. 80, pp. 14–23, 2015, doi: 10.1016/j.knosys.2015.01.010.
[69] A. P. Engelbrecht, *Computational Intelligence: An Introduction: Second Edition*. 2007.
[70] L. Rutkowski, *Computational intelligence: Methods and techniques*. 2008.
[71] J. S. Raj, "A COMPREHENSIVE SURVEY ON THE COMPUTATIONAL INTELLIGENCE TECHNIQUES AND ITS APPLICATIONS," *J. ISMAC*, vol. 01, no. 03, 2019, doi: 10.36548/jismac.2019.3.002.
[72] J. C. Bezdek, "(Computational) Intelligence: What's in a Name?," *IEEE Syst. Man, Cybern. Mag.*, vol. 2, no. 2, 2016, doi: 10.1109/msmc.2016.2558778.
[73] B. B. Gupta, A. K. Sangaiah, N. Nedjah, S. Yamaguchi, Z. Zhang, and M. Sheng, "Recent research in computational intelligence paradigms into security and privacy for online social networks (OSNs)," *Future Generation Computer Systems*, vol. 86. 2018, doi: 10.1016/j.future.2018.05.017.
[74] R. Iqbal, F. Doctor, B. More, S. Mahmud, and U. Yousuf, "Big data analytics: Computational intelligence techniques and application areas," *Technol. Forecast. Soc. Change*, vol. 153, 2020, doi: 10.1016/j.techfore.2018.03.024.
[75] H. M. Said and A. B. M. Salem, "Exploiting Computational Intelligence Paradigms in e-Technologies and Activities," in *Procedia Computer Science*, 2015, vol. 65, doi: 10.1016/j.procs.2015.09.101.
[76] M. Vukobratović, P. Marić, G. Horvat, Z. Balkić, and S. Sučić, "A survey on computational intelligence applications in distribution network optimization," *Electronics (Switzerland)*, vol. 10, no. 11. 2021, doi: 10.3390/electronics10111247.
[77] E. Gomede, F. H. Gaffo, G. U. Briganó, R. M. de Barros, and L. de S. Mendes, "Application of computational intelligence to improve education in smart cities," *Sensors (Switzerland)*, vol. 18, no. 1, 2018, doi: 10.3390/s18010267.
[78] R. M. De Moraes and L. Martínez, "Computational Intelligence Applications for Data Science," *Knowledge-Based Systems*, vol. 87. pp. 1–2, 2015, doi: 10.1016/j.knosys.2015.07.038.
[79] L. A. Zadeh, "Fuzzy logic," *Computer (Long. Beach. Calif).*, 1988, doi: 10.1109/2.53.
[80] M. H. Hassoun, "Fundamentals of Artificial Neural Networks," *Proc. IEEE*, vol. 84,

no. 6, 2005, doi: 10.1109/jproc.1996.503146.
[81] S. Haykin, *Neural Networks and Learning Machines*, vol. 3. 2008.
[82] C. A. Coello, “An introduction to evolutionary algorithms and their applications,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2005, vol. 3563 LNCS, doi: 10.1007/11533962_39.
[83] S. A. Hofmeyr and S. Forrest, “Architecture for an artificial immune system.,” *Evol. Comput.*, vol. 8, no. 4, 2000, doi: 10.1162/106365600568257.
[84] B. Ruhnau, “Eigenvector-centrality - a node-centrality,” *Soc. Networks*, 2000, doi: 10.1016/S0378-8733(00)00031-9.
[85] T. Opsahl, F. Agneessens, and J. Skvoretz, “Node centrality in weighted networks: Generalizing degree and shortest paths,” *Soc. Networks*, vol. 32, no. 3, pp. 245–251, 2010, doi: 10.1016/j.socnet.2010.03.006.
[86] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, “Complex networks: Structure and dynamics,” *Physics Reports*. 2006, doi: 10.1016/j.physrep.2005.10.009.
[87] L. C. Freeman, “Centrality in Social Networks,” *Soc. Networks*, vol. 1, no. 3, pp. 215–239, 1978, doi: 10.1016/0378-8733(78)90021-7.
[88] S. Al-Oufi, H. N. Kim, and A. El Saddik, “A group trust metric for identifying people of trust in online social networks,” *Expert Syst. Appl.*, vol. 39, no. 18, pp. 13173–13181, 2012, doi: 10.1016/j.eswa.2012.05.084.
[89] F. Bodendorf and C. Kaiser, “Detecting opinion leaders and trends in online communities,” *4th Int. Conf. Digit. Soc. ICDS 2010, Incl. CYBERLAWS 2010 1st Int. Conf. Tech. Leg. Asp. e-Society*, pp. 124–129, 2010, doi: 10.1109/ICDS.2010.29.
[90] M. F. Tsai, C. W. Tzeng, Z. L. Lin, and A. L. P. Chen, “Discovering leaders from social network by action cascade,” *Soc. Netw. Anal. Min.*, vol. 4, no. 1, pp. 1–10, 2014, doi: 10.1007/s13278-014-0165-9.
[91] Q. Miao, S. Zhang, Y. Meng, and H. Yu, “Domain-sensitive opinion leader mining from online review communities,” in *Proceedings of the 22nd International Conference on World Wide Web - WWW ’13 Companion*, 2013, pp. 187–188, doi: 10.1145/2487788.2487882.
[92] Y. Cho, J. Wang, and D. Lee, “Identification of effective opinion leaders in the diffusion of technological innovation: A social network approach,” *Technol. Forecast. Soc. Change*, vol. 79, no. 1, pp. 97–106, 2012, doi: 10.1016/j.techfore.2011.06.003.
[93] J. Duan, J. Zeng, and B. Luo, “Identification of opinion leaders based on user clustering and sentiment analysis,” in *Proceedings - 2014 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IAT 2014*, 2014, pp. 377–383, doi: 10.1109/WI-IAT.2014.59.
[94] M. Z. Shafiq, M. U. Ilyas, A. X. Liu, and H. Radha, “Identifying leaders and followers in online social networks,” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9,

pp. 618–628, 2013, doi: 10.1109/JSAC.2013.SUP.0513054.
[95] A. Aleahmad, P. Karisani, M. Rahgozar, and F. Oroumchian, "OLFinder: Finding opinion leaders in online social networks," *J. Inf. Sci.*, vol. 42, no. 5, pp. 656–674, 2016, doi: 10.1177/0165551515605217.
[96] Y. Chen, L. Hui, C. I. Wu, H. Liu, and S. Chen, "Opinion Leaders Discovery in Dynamic Social Network," 2017.
[97] S. M. Aghdam and N. Jafari Navimipour, "Opinion leaders selection in the social networks based on trust relationships propagation," *Karbala Int. J. Mod. Sci.*, vol. 2, no. 2, pp. 88–97, 2016, doi: 10.1016/j.kijoms.2016.02.002.
[98] B. Huang, G. Yu, and H. R. Karimi, "The finding and dynamic detection of opinion leaders in social network," *Math. Probl. Eng.*, pp. 1–8, 2014, doi: 10.1155/2014/328407.
[99] Y. C. Chen, Y. H. Chen, C. H. Hsu, H. J. You, J. Liu, and X. Huang, "Mining opinion leaders in big social network," in *Proceedings - International Conference on Advanced Information Networking and Applications, AINA*, 2017, pp. 1012–1018, doi: 10.1109/AINA.2017.147.
[100] Y. Liu, Z. Gu, T. H. Ko, and J. Liu, "Identifying Key Opinion Leaders in Social Media via Modality-Consistent Harmonized Discriminant Embedding," *IEEE Trans. Cybern.*, pp. 1–12, 2018, doi: 10.1109/TCYB.2018.2871765.
[101] G. Sun and S. Bin, "A new opinion leaders detecting algorithm in multi-relationship online social networks," *Multimed. Tools Appl.*, vol. 77, no. 4, pp. 4295–4307, 2018, doi: 10.1007/s11042-017-4766-y.
[102] L. Yang, Y. Qiao, Z. Liu, J. Ma, and X. Li, "Identifying opinion leader nodes in online social networks with a new closeness evaluation algorithm," *Soft Comput.*, vol. 22, no. 2, pp. 453–464, 2018, doi: 10.1007/s00500-016-2335-3.
[103] L. Yang, Y. Tian, J. Li, J. Ma, and J. Zhang, "Identifying opinion leaders in social networks with topic limitation," *Cluster Comput.*, vol. 20, no. 3, pp. 2403–2413, 2017, doi: 10.1007/s10586-017-0732-8.
[104] S. Walter and M. Brüggemann, "Opportunity makes opinion leaders: analyzing the role of first-hand information in opinion leadership in social media networks," *Inf. Commun. Soc.*, 2020, doi: 10.1080/1369118X.2018.1500622.
[105] Z. Wang, H. Liu, W. Liu, and S. Wang, "Understanding the power of opinion leaders' influence on the diffusion process of popular mobile games: Travel Frog on Sina Weibo," *Comput. Human Behav.*, 2020, doi: 10.1016/j.chb.2020.106354.
[106] Z. Bu, H. J. Li, C. Zhang, J. Cao, A. Li, and Y. Shi, "Graph K-means Based on Leader Identification, Dynamic Game, and Opinion Dynamics," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 7, pp. 1348–1361, 2020, doi: 10.1109/TKDE.2019.2903712.
[107] J. Cao, Z. Bu, Y. Wang, H. Yang, J. Jiang, and H. J. Li, "Detecting Prosumer-Community Groups in Smart Grids from the Multiagent Perspective," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2019, doi: 10.1109/TSMC.2019.2899366.
[108] F. Bodendorf and C. Kaiser, "Detecting Opinion Leaders and Trends in Online

Social Networks," *Proc. 2nd ACM Work. Soc. web search Min.*, 2009, doi: 10.1109/ICDS.2010.29.

[109] Y. C. Chen, "A novel algorithm for mining opinion leaders in social networks," *World Wide Web*, vol. 22, no. 3, pp. 1279–1295, 2019, doi: 10.1007/s11280-018-0586-x.

[110] F. K. Dewi, S. B. Yudhoatmojo, and I. Budi, "Identification of opinion leader on rumor spreading in online social network Twitter using edge weighting and centrality measure weighting," 2018, doi: 10.1109/ICDIM.2017.8244680.

[111] Q. Liqing, D. Jinlong, L. Haiyan, and W. Yan, "Detecting opinion leaders in online social networks using HybridRank algorithm," *J. Intell. Fuzzy Syst.*, vol. 35, no. 1, pp. 513–522, 2018, doi: 10.3233/JIFS-169607.

[112] F. Riquelme, P. Gonzalez-Cantergiani, D. Hans, R. Villarroel, and R. Munoz, "Identifying Opinion Leaders on Social Networks Through Milestones Definition," *IEEE Access*, vol. 7, pp. 75670–75677, 2019, doi: 10.1109/ACCESS.2019.2922155.

[113] S. Lei, X. Zhang, and Y. Yan, "Research on opinion leaders recognition based on TOPSIS in open source design community," in *Proceedings of the 2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design, CSCWD 2019*, 2019, pp. 307–312, doi: 10.1109/CSCWD.2019.8791850.

[114] J. Yang, Y. Zhang, and L. Liu, "Identifying Opinion Leaders in Virtual Travel Community Based on Social Network Analysis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11589 LNCS, pp. 276–294, doi: 10.1007/978-3-030-22338-0_23.

[115] A. U. Rehman, A. Jiang, A. Rehman, A. Paul, S. Din, and M. T. Sadiq, "Identification and role of opinion leaders in information diffusion for online discussion network," *J. Ambient Intell. Humaniz. Comput.*, Jan. 2020, doi: 10.1007/s12652-019-01623-5.

[116] C. Xia, S. Sun, F. Rao, J. Sun, J. Wang, and Z. Chen, "SIS model of epidemic spreading on dynamical networks with community," *Front. Comput. Sci. China*, vol. 3, no. 3, pp. 361–365, 2009, doi: 10.1007/s11704-009-0057-8.

[117] Y. Zhang and J. Xu, "A rumor spreading model considering the cumulative effects of memory," *Discret. Dyn. Nat. Soc.*, vol. 2015, pp. 1–11, 2015, doi: 10.1155/2015/204395.

[118] N. Zhang, H. Huang, B. Su, J. Zhao, and B. Zhang, "Dynamic 8-state ICSAR rumor propagation model considering official rumor refutation," *Phys. A Stat. Mech. its Appl.*, vol. 415, pp. 333–346, Dec. 2014, doi: 10.1016/j.physa.2014.07.023.

[119] Y. Q. Wang, X. Y. Yang, Y. L. Han, and X. A. Wang, "Rumor spreading model with trust mechanism in complex social networks," *Commun. Theor. Phys.*, vol. 59, no. 4, pp. 510–516, 2013, doi: 10.1088/0253-6102/59/4/21.

[120] A. Bodaghi and S. Goliaei, "A NOVEL MODEL for RUMOR SPREADING on SOCIAL NETWORKS with CONSIDERING the INFLUENCE of DISSENTING OPINIONS," *Adv. Complex Syst.*, vol. 21, no. 6–7, 2018, doi: 10.1142/S021952591850011X.

[121] Y. Hu, Q. Pan, W. Hou, and M. He, "Rumor spreading model with the different attitudes towards rumors," *Phys. A Stat. Mech. its Appl.*, vol. 502, pp. 331–344, Jul. 2018, doi: 10.1016/j.physa.2018.02.096.

[122] J. J. Cheng, Y. Liu, B. Shen, and W. G. Yuan, "An epidemic model of rumor diffusion in online social networks," *Eur. Phys. J. B*, vol. 86, no. 1, p. 29, Jan. 2013, doi: 10.1140/epjb/e2012-30483-5.

[123] P. Jia, C. Wang, G. Zhang, and J. Ma, "A rumor spreading model based on two propagation channels in social networks," *Phys. A Stat. Mech. its Appl.*, vol. 524, pp. 342–353, Jun. 2019, doi: 10.1016/j.physa.2019.04.163.

[124] L. Zhu and X. Huang, "SIS Model of Rumor Spreading in Social Network with Time Delay and Nonlinear Functions," *Commun. Theor. Phys.*, vol. 72, no. 1, p. 015002, Jan. 2020, doi: 10.1088/1572-9494/ab4ef6.

[125] Y. Xiao, D. Chen, S. Wei, Q. Li, H. Wang, and M. Xu, "Rumor propagation dynamic model based on evolutionary game and anti-rumor," *Nonlinear Dyn.*, vol. 95, no. 1, pp. 523–539, 2019, doi: 10.1007/s11071-018-4579-1.

[126] X. S. Yang and X. He, "Firefly algorithm: recent advances and applications," *Int. J. Swarm Intell.*, 2013, doi: 10.1504/IJSI.2013.055801.

[127] L. Sthle and S. Wold, "Analysis of variance (ANOVA)," *Chemometrics and Intelligent Laboratory Systems*. 1989, doi: 10.1016/0169-7439(89)80095-4.

[128] I. Fister, X. S. Yang, and J. Brest, "A comprehensive review of firefly algorithms," *Swarm Evol. Comput.*, 2013, doi: 10.1016/j.swevo.2013.06.001.

[129] F. Dabaghi-Zarandi and M. Kuchaki Rafsanjani, "Community Detection in Social Networks," in *Studies in Systems, Decision and Control*, 2019, pp. 273–293.

[130] D. Mehrle, A. Strosser, and A. Harkin, "Walk-modularity and community structure in networks," *Netw. Sci.*, vol. 3, no. 3, pp. 348–360, 2015, doi: 10.1017/nws.2015.20.

[131] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Phys. Rep.*, vol. 659, pp. 1–44, 2016, doi: 10.1016/j.physrep.2016.09.002.

[132] M. E. J. Newman and T. P. Peixoto, "Generalized Communities in Networks," *Phys. Rev. Lett.*, vol. 115, no. 8, p. 088701, 2015, doi: 10.1103/PhysRevLett.115.088701.

[133] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, "Generalized Louvain method for community detection in large networks," in *International Conference on Intelligent Systems Design and Applications, ISDA*, 2011, pp. 88–93, doi: 10.1109/ISDA.2011.6121636.

[134] S. Papadopoulos, Y. Kompatsiaris, A. Vakali, and P. Spyridonos, "Community detection in Social Media," *Data Min. Knowl. Discov.*, vol. 24, no. 3, pp. 515–554, 2012, doi: 10.1007/s10618-011-0224-z.

[135] P. Chejara and W. W. Godfrey, "Comparative analysis of community detection algorithms," in *2017 Conference on Information and Communication Technology, CICT 2017*, 2018, pp. 1–5, doi: 10.1109/INFOCOMTECH.2017.8340627.

[136] L. Hou and K. Liu, "Common neighbour structure and similarity intensity in complex networks," *Phys. Lett. Sect. A Gen. At. Solid State Phys.*, vol. 381, no. 39, 2017, doi: 10.1016/j.physleta.2017.08.050.

[137] J. Cheng *et al.*, “Neighbor Similarity Based Agglomerative Method for Community Detection in Networks,” *Complexity*, vol. 2019, 2019, doi: 10.1155/2019/8292485.
[138] “Unsupervised learning: Foundations of neural computation,” *Comput. Math. with Appl.*, vol. 38, no. 5–6, 1999, doi: 10.1016/s0898-1221(99)90165-7.
[139] Michael J. Garbade, M. Garbade, and Michael J. Garbade, “Understanding K-means Clustering in Machine Learning,” *Towards Data Science*, 2018. .
[140] M. Kwan and D. Ramachandran, “Trust and Online Reputation Systems,” 2009, pp. 287–311.
[141] L. Aroyo, P. De Meo, and D. Ursino, “Trust and reputation in social internetworking systems,” in *CEUR Workshop Proceedings*, 2010, p. 42.
[142] M. El Marrakchi, M. Bellafkih, and H. Bensaid, “Towards reputation measurement in online social networks,” in *2015 Intelligent Systems and Computer Vision, ISCV 2015*, 2015, pp. 1–8, doi: 10.1109/ISACV.2015.7105540.
[143] G. Swamynathan, K. C. Almeroth, and B. Y. Zhao, “The design of a reliable reputation system,” *Electron. Commer. Res.*, vol. 10, no. 3–4, pp. 239–270, 2010, doi: 10.1007/s10660-010-9064-y.
[144] M. McPherson, L. Smith-Lovin, and J. M. Cook, “Birds of a Feather: Homophily in Social Networks,” *Annu. Rev. Sociol.*, vol. 27, no. 1, pp. 415–444, 2001, doi: 10.1146/annurev.soc.27.1.415.
[145] S. Mirjalili and A. Lewis, “The Whale Optimization Algorithm,” *Adv. Eng. Softw.*, vol. 95, pp. 51–67, 2016, doi: 10.1016/j.advengsoft.2016.01.008.
[146] M. Mafarja and S. Mirjalili, “Whale optimization approaches for wrapper feature selection,” *Appl. Soft Comput. J.*, vol. 62, pp. 441–453, 2018, doi: 10.1016/j.asoc.2017.11.006.
[147] M. M. Mafarja and S. Mirjalili, “Hybrid Whale Optimization Algorithm with simulated annealing for feature selection,” *Neurocomputing*, vol. 260, pp. 302–312, 2017, doi: 10.1016/j.neucom.2017.04.053.
[148] Q. V. Pham, S. Mirjalili, N. Kumar, M. Alazab, and W. J. Hwang, “Whale Optimization Algorithm with Applications to Resource Allocation in Wireless Networks,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, 2020, doi: 10.1109/TVT.2020.2973294.
[149] Y. Yue, H. You, S. Wang, and L. Cao, “Improved whale optimization algorithm and its application in heterogeneous wireless sensor networks,” *Int. J. Distrib. Sens. Networks*, vol. 17, no. 5, 2021, doi: 10.1177/15501477211018140.
[150] Z. Yan, S. Wang, B. Liu, and X. Li, “Application of Whale Optimization Algorithm in Optimal Allocation of Water Resources,” in *E3S Web of Conferences*, 2018, vol. 53, doi: 10.1051/e3sconf/20185304019.
[151] K. C. Lee and P. T. Lu, “Application of Whale Optimization Algorithm to Inverse Scattering of an Imperfect Conductor with Corners,” *Int. J. Antennas Propag.*, vol. 2020, 2020, doi: 10.1155/2020/8205797.
[152] N. Rana, M. S. A. Latiff, S. M. Abdulhamid, and H. Chiroma, “Whale optimization algorithm: a systematic review of contemporary applications, modifications and developments,” *Neural Computing and Applications*, vol. 32, no. 20. pp. 16245–

16277, 2020, doi: 10.1007/s00521-020-04849-z.

[153] H. Chen, Y. Xu, M. Wang, and X. Zhao, "A balanced whale optimization algorithm for constrained engineering design problems," *Appl. Math. Model.*, vol. 71, pp. 45–59, 2019, doi: 10.1016/j.apm.2019.02.004.

[154] M. A. El Aziz, A. A. Ewees, and A. E. Hassanien, "Whale Optimization Algorithm and Moth-Flame Optimization for multilevel thresholding image segmentation," *Expert Syst. Appl.*, vol. 83, pp. 242–256, 2017, doi: 10.1016/j.eswa.2017.04.023.

[155] L. C. Freeman, "Centrality in social networks conceptual clarification," *Soc. Networks*, vol. 1, no. 3, pp. 215–239, 1978, doi: 10.1016/0378-8733(78)90021-7.

[156] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting Positive and Negative Links," in *International World Wide Web Conference*, 2010, pp. 641–650, doi: 10.1145/1772690.1772756.

[157] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," 2007, pp. 29–42, doi: 10.1145/1298306.1298311.

[158] R. G. Regis and C. A. Shoemaker, "Local function approximation in evolutionary algorithms for the optimization of costly functions," *IEEE Trans. Evol. Comput.*, vol. 8, no. 5, pp. 490–505, 2004, doi: 10.1109/TEVC.2004.835247.

[159] M. Dorigo and G. Di Caro, "Ant colony optimization: a new meta-heuristic," *Proc. 1999 Congr. Evol. Comput. (Cat. No. 99TH8406)*, vol. 2, pp. 1470–1477, 1999, doi: 10.1109/CEC.1999.782657.

[160] X.-S. Yang, *Nature-Inspired Metaheuristic Algorithms*. 2010.

[161] S. Sanadhya and S. Singh, "Trust Calculation with Ant Colony Optimization in Online Social Networks," *Procedia Comput. Sci.*, vol. 54, pp. 186–195, 2015, doi: 10.1016/j.procs.2015.06.021.

[162] P. P. Shenoy, "On coalition formation: a game-theoretical approach," *Int. J. Game Theory*, 1979, doi: 10.1007/BF01770064.

[163] D. Ray, *A Game-Theoretic Perspective on Coalition Formation*. 2007.

[164] Y. Narahari and R. Narayanam, "Tutorial: Game theoretic models for social network analysis," 2011, doi: 10.1145/1963192.1963316.

[165] M. del Pozo, C. Manuel, E. González-Arangüena, and G. Owen, "Centrality in directed social networks. A game theoretic approach," *Soc. Networks*, 2011, doi: 10.1016/j.socnet.2011.04.001.

[166] M. H. Manshaei, Q. Zhu, T. Alpcan, T. Basar, and J. P. Hubaux, "Game theory meets network security and privacy," *ACM Computing Surveys*. 2013, doi: 10.1145/2480741.2480742.

[167] S. Zhang and Y. Zhang, "Introduction to game theory," *Chinese Science Bulletin*. 2003, doi: 10.1360/99ws0111.

[168] R. Leonard, "Game theory," in *Handbook on the History of Economic Analysis*, 2016.

[169] W. P. Fox and R. Burks, "Game Theory," in *International Series in Operations Research and Management Science*, 2019.

[170] K. Leyton-Brown and Y. Shoham, "Essentials of Game Theory: A Concise

Multidisciplinary Introduction," *Synth. Lect. Artif. Intell. Mach. Learn.*, 2008, doi: 10.2200/s00108ed1v01y200802aim003.

[171] C. F. Camerer, *Behavioral game theory: Experiments in strategic interaction*. 2003.

[172] J. von Neumann and O. Morgenstern, *Theory of games and economic behavior*. 2007.

[173] W. Thomson and A. E. Roth, "The Shapley Value: Essays in Honor of Lloyd S. Shapley.," *Economica*, 1991, doi: 10.2307/2554979.

[174] L. S. Shapley, "17. A Value for n-Person Games," in *Contributions to the Theory of Games (AM-28), Volume II*, 2016.

[175] T. Ui, "A Shapley Value Representation of Potential Games," *Games Econ. Behav.*, 2000, doi: 10.1006/game.1999.0729.

[176] R. Narayanam and Y. Narahari, "A shapley value-based approach to discover influential nodes in social networks," *IEEE Trans. Autom. Sci. Eng.*, 2011, doi: 10.1109/TASE.2010.2052042.

[177] T. P. Michalak, K. V. Aadithya, P. L. Szczepanski, B. Ravindran, and N. R. Jennings, "Efficient computation of the shapley value for game-theoretic network centrality," *J. Artif. Intell. Res.*, 2013, doi: 10.1613/jair.3806.

[178] E. Hernandez and J. M. Shaver, "Network Synergy," *Adm. Sci. Q.*, 2019, doi: 10.1177/0001839218761369.

[179] V. Buskens, "The social structure of trust," *Soc. Networks*, 1998, doi: 10.1016/S0378-8733(98)00005-7.

[180] L. Liu and W. Shi, "Trust and reputation management," *IEEE Internet Comput.*, 2010, doi: 10.1109/MIC.2010.124.

[181] B. A. Brooks, "The Strength of Weak Ties," *Nurse Lead.*, 2019, doi: 10.1016/j.mnl.2018.12.011.

[182] M. S. Granovetter, "The Strength of Weak Ties," *Am. J. Sociol.*, vol. 78, no. 6, pp. 1360–1380, 1973, doi: 10.1086/225469.

[183] R. S. Burt, "Structural holes and good ideas," *Am. J. Sociol.*, 2004, doi: 10.1086/421787.

[184] S. Goyal and F. Vega-Redondo, "Structural holes in social networks," *J. Econ. Theory*, 2007, doi: 10.1016/j.jet.2007.01.006.

[185] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks," *Third Int. AAAI Conf. Weblogs Soc. Media*, 2009, doi: 10.1136/qshc.2004.010033.

[186] M. F. Sanner, "Python: A programming language for software integration and development," *Journal of Molecular Graphics and Modelling*. 1999.

[187] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," 2010, doi: 10.1145/1772690.1772756.

[188] S. Kumar, F. Spezzano, V. S. Subrahmanian, and C. Faloutsos, "Edge weight prediction in weighted signed networks," 2017, doi: 10.1109/ICDM.2016.175.

[189] Ministerial Meeting on Population of the Non-Aligned Movement (1993: Bali), "Denpasar Declaration on Population and Development.," *Integration*, no. 40, pp. 27–9, Nov. 1994, doi: 10.1234/2013/999990.

[190] P. S. Van Eck, W. Jager, and P. S. H. Leeflang, "Opinion leaders' role in innovation diffusion: A simulation study," *J. Prod. Innov. Manag.*, 2011, doi: 10.1111/j.1540-5885.2011.00791.x.

[191] R. Cressman and J. Apaloo, "Evolutionary game theory," in *Handbook of Dynamic Game Theory*, 2018.

[192] "Computational intelligence: a logical approach," *Choice Rev. Online*, 1998, doi: 10.5860/choice.35-5701.

[193] W. G. Hatcher and W. Yu, "A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends," *IEEE Access*, vol. 6, 2018, doi: 10.1109/ACCESS.2018.2830661.

[194] L. Deng and D. Yu, "Deep learning: Methods and applications," *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4. 2013, doi: 10.1561/2000000039.

[195] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *J. Big Data*, vol. 2, no. 1, 2015, doi: 10.1186/s40537-014-0007-7.

[196] J. Ahmad, H. Farman, and Z. Jan, "Deep Learning Methods and Applications," in *SpringerBriefs in Computer Science*, 2019.

[197] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Transactions on Signal and Information Processing*, vol. 3. 2014, doi: 10.1017/ATSIP.2013.99.

[198] M. Z. Alom *et al.*, "A state-of-the-art survey on deep learning theory and architectures," *Electronics (Switzerland)*, vol. 8, no. 3. 2019, doi: 10.3390/electronics8030292.

[199] J. Suárez-Varela *et al.*, "Challenging the generalization capabilities of graph neural networks for network modeling," 2019, doi: 10.1145/3342280.3342327.

[200] P. Goyal and E. Ferrara, "Graph embedding techniques, applications, and performance: A survey," *Knowledge-Based Syst.*, vol. 151, 2018, doi: 10.1016/j.knosys.2018.03.022.

[201] H. Steeve, "A Gentle Introduction to Graph Neural Networks (Basics, DeepWalk, and GraphSage)," *Https://Medium.Com/@Huangkh19951228*, 2019.

[202] J. Qiu, Y. Dong, H. Ma, J. Li, K. Wang, and J. Tang, "Network embedding as matrix factorization: Unifying DeepWalk, LINE, PTE, and node2vec," in *WSDM 2018 - Proceedings of the 11th ACM International Conference on Web Search and Data Mining*, 2018, vol. 2018-Febuary, doi: 10.1145/3159652.3159706.

[203] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: Online learning of social representations," 2014, doi: 10.1145/2623330.2623732.

[204] C. C. Kiliroor and C. Valliyammai, "Trust analysis on social networks for identifying authenticated users," *2016 8th Int. Conf. Adv. Comput. ICoAC 2016*, pp. 37–41, 2017, doi: 10.1109/ICoAC.2017.7951741.

[205] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, "Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters," *Internet Math.*, vol. 6, no. 1, 2009, doi: 10.1080/15427951.2009.10129177.

[206] M. Richardson, R. Agrawal, and P. Domingos, “Trust management for the Semantic Web,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2870, 2003, doi: 10.1007/978-3-540-39718-2_23.

[207] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, “Group formation in large social networks: Membership, growth, and evolution,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006, vol. 2006.

[208] T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, and P. Lamere, “The million song dataset,” 2011.

[209] J. Zhang, B. Liu, J. Tang, T. Chen, and J. Li, “Social influence locality for modeling retweeting behaviors,” 2013.

[210] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.

[211] D. M. W. POWERS, “Denpasar Declaration on Population and Development.,” *Integration*, no. 40, pp. 27–29, 1994, doi: 10.1234/2013/999990.

[212] P. Thagard, “Machine Learning,” in *A Companion to Cognitive Science*, 2008.

[213] N. P. Nguyen, T. N. Dinh, Y. Xuan, and M. T. Thai, “Adaptive algorithms for detecting community structure in dynamic social networks,” 2011, doi: 10.1109/INFCOM.2011.5935045.

[214] V. Sekara, A. Stopczynski, and S. Lehmann, “Fundamental structures of dynamic social networks,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 36, 2016, doi: 10.1073/pnas.1602803113.

[215] World Health Organization, “WHO characterizes COVID-19 as a pandemic,” *World Health Organization*, 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

[216] R. Radu, “Fighting the ‘Infodemic’: Legal Responses to COVID-19 Disinformation,” *Soc. Media + Soc.*, vol. 6, no. 3, p. 205630512094819, Jul. 2020, doi: 10.1177/2056305120948190.

[217] M. Palacios Cruz, E. Santos, M. A. Velázquez Cervantes, and M. León Juárez, “COVID-19, una emergencia de salud pública mundial,” *Rev. Clínica Española*, vol. 221, no. 1, pp. 55–61, Jan. 2021, doi: 10.1016/j.rce.2020.03.001.

[218] M. S. Islam *et al.*, “COVID-19-Related infodemic and its impact on public health: A global social media analysis,” *Am. J. Trop. Med. Hyg.*, vol. 103, no. 4, pp. 1621–1629, 2020, doi: 10.4269/ajtmh.20-0812.

[219] L. P. Garcia and E. Duarte, “Infodemic: excess quantity to the detriment of quality of information about COVID-19.,” *Epidemiol. e Serv. saude Rev. do Sist. Unico Saude do Bras.*, vol. 29, no. 4, 2020, doi: 10.1590/S1679-49742020000400019.

[220] M. Alshible, “Social media rumors in time of corona pandemic, why & how is criminalized? (comparative study),” *J. Adv. Res. Law Econ.*, vol. 11, no. 3, pp. 719–725, 2020, doi: 10.14505/jarle.v11.3(49).03.

[221] D. H. Shin, “The effects of trust, security and privacy in social networking: A security-based approach to understand the pattern of adoption,” *Interact. Comput.*, vol. 22, no. 5, pp. 428–438, 2010, doi: 10.1016/j.intcom.2010.05.001.

[222] B. Doerr, M. Fouz, and T. Friedrich, "Why rumors spread so quickly in social networks," *Commun. ACM*, vol. 55, no. 6, pp. 70–75, 2012, doi: 10.1145/2184319.2184338.

[223] H. Hui, C. Zhou, X. Lü, and J. Li, "Spread mechanism and control strategy of social network rumors under the influence of COVID-19," *Nonlinear Dyn.*, vol. 101, no. 3, pp. 1933–1949, 2020, doi: 10.1007/s11071-020-05842-w.

[224] D. H. Zanette, "Dynamics of rumor propagation on small-world networks," *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.*, vol. 65, no. 4, 2002, doi: 10.1103/PhysRevE.65.041908.

[225] A. Abdoli, "Gossip, rumors and the COVID-19 crisis," *Disaster Medicine and Public Health Preparedness*. 2020, doi: 10.1017/dmp.2020.272.

[226] S. H. Saadat *et al.*, "Fear and Panic of COVID-19," *Int. J. Travel Med. Glob. Heal.*, vol. 8, no. 3, 2020, doi: 10.34172/ijtmgh.2020.15.

[227] N. Keys, D. C. Thomsen, and T. F. Smith, "Opinion leaders and complex sustainability issues," *Manag. Environ. Qual. An Int. J.*, vol. 21, no. 2, pp. 187–197, 2010, doi: 10.1108/14777831011025535.

[228] J.-X. Zhang, D.-B. Chen, Q. Dong, and Z.-D. Zhao, "Identifying a set of influential spreaders in complex networks," *Sci. Rep.*, vol. 6, no. 1, p. 27823, Jun. 2016, doi: 10.1038/srep27823.

[229] L. Jain and R. Katarya, "A Systematic Survey of Opinion Leader in Online Social Network," in *ICSNS 2018 - Proceedings of IEEE International Conference on Soft-Computing and Network Security*, Feb. 2018, pp. 1–5, doi: 10.1109/ICSNS.2018.8573639.

[230] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, 2014, pp. 216–225.

[231] J. Golbeck, "Trust and Nuanced Profile Similarity in Online Social Networks," *ACM Trans. Web*, vol. 3, no. 4, pp. 1–33, 2009, doi: 10.1145/1594173.1594174.

[232] L. Jain and R. Katarya, "Identification of opinion leader in online social network using fuzzy trust system," in *2018 IEEE 8th International Advance Computing Conference (IACC)*, Dec. 2018, pp. 233–239, doi: 10.1109/IADCC.2018.8692095.

[233] P. De Meo, E. Ferrara, D. Rosaci, and G. M. L. Sarné, "Trust and compactness in social network groups," *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 205–216, 2015, doi: 10.1109/TCYB.2014.2323892.

[234] E. Dichter, "How Word of Mouth Advertising Works," *Harv. Bus. Rev.*, 1966, doi: 10.1016/S0267-3649(00)88914-1.

[235] J. Roozenbeek *et al.*, "Susceptibility to misinformation about COVID-19 around the world: Susceptibility to COVID misinformation," *R. Soc. Open Sci.*, vol. 7, no. 10, p. 201199, Oct. 2020, doi: 10.1098/rsos.201199.

[236] R. Kouzy *et al.*, "Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter," *Cureus*, Mar. 2020, doi: 10.7759/cureus.7255.

[237] C. Zhou, H. Xiu, Y. Wang, and X. Yu, "Characterizing the dissemination of

misinformation on social media in health emergencies: An empirical study based on COVID-19," *Inf. Process. Manag.*, vol. 58, no. 4, p. 102554, Jul. 2021, doi: 10.1016/j.ipm.2021.102554.

[238] B. Skyrms and R. Pemantle, "A dynamic model of social network formation," *Proc. Natl. Acad. Sci. U. S. A.*, 2000, doi: 10.1073/pnas.97.16.9340.

[239] I. Fister, X. S. Yang, J. Brest, and D. Fister, "A brief review of nature-inspired algorithms for optimization," *Elektrotehniski Vestnik/Electrotechnical Review*, vol. 80, no. 3. 2013.

[240] Q. Tan, N. Liu, and X. Hu, "Deep representation learning for social network analysis," *arXiv*. 2019, doi: 10.3389/fdata.2019.00002.

[241] S. Cui, T. Li, S. C. Chen, M. L. Shyu, Q. Li, and H. Zhang, "DISL: Deep Isomorphic Substructure Learning for network representations," *Knowledge-Based Syst.*, vol. 189, 2020, doi: 10.1016/j.knosys.2019.105086.

[242] F. Yin, X. Xia, N. Song, L. Zhu, and J. Wu, "Quantify the role of superspreaders - opinion leaders- On COVID-19 information propagation in the Chinese Sina-microblog," *PLoS One*, vol. 15, no. 6, 2020, doi: 10.1371/journal.pone.0234023.

[243] L. Weng and Q. Zhang, "A social recommendation method based on opinion leaders," *Multimed. Tools Appl.*, vol. 80, no. 4, 2021, doi: 10.1007/s11042-020-09972-6.

[244] S. Tobon and J. García-Madariaga, "The influence of opinion leaders' ewom on online consumer decisions: A study on social influence," *J. Theor. Appl. Electron. Commer. Res.*, vol. 16, no. 4, 2021, doi: 10.3390/jtaer16040043.

[245] Š. Žák and M. Hasprová, "The Impact of Opinion Leaders on the Consumer Behaviour in the Global Digital Environment," *SHS Web Conf.*, vol. 92, 2021, doi: 10.1051/shsconf/20219206043.

[246] C. Torrens, F. Santana, and J. Miquel, "Opinion leaders in the medical community attract more attention than randomized controlled trials in shoulder surgery," *Int. Orthop.*, vol. 45, no. 1, 2021, doi: 10.1007/s00264-020-04775-8.

[247] N. Ghali, M. Panda, A. E. Hassanien, A. Abraham, and V. Snasel, "Social networks analysis: Tools, measures and visualization," in *Computational Social Networks: Mining and Visualization*, vol. 9781447140542, 2012.

[248] R. Katarya and O. P. Verma, "Recent developments in affective recommender systems," *Phys. A Stat. Mech. its Appl.*, 2016, doi: 10.1016/j.physa.2016.05.046.

[249] A. Zare Ravasan, S. Rouhani, and S. Asgary, "A Review for the Online Social Networks Literature (2005-2011)," *Eur. J. Bus. Manag. www.iiste.org ISSN*, vol. 6, no. 4, 2014.

[250] S. Al-Yazidi, J. Berri, M. Al-Qurishi, and M. Al-Alrubaian, "Measuring Reputation and Influence in Online Social Networks: A Systematic Literature Review," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2999033.

[251] M. Shi and A. C. Wojnicki, "Money talks...to online opinion leaders: What motivates opinion leaders to make social-network referrals?," *J. Advert. Res.*, vol. 54, no. 1, 2014, doi: 10.2501/JAR-54-1-081-091.

List of publications

Journals

1. Jain, L., & Katarya, R. (2019). Discover opinion leader in online social network using firefly algorithm. Expert Systems with Applications, 122, 1-15. **[Published, SCIE, IF:5.452]**
2. Jain, L., Katarya, R., & Sachdeva, S. (2020). Opinion leader detection using whale optimization algorithm in online social network. Expert Systems with Applications, 142, 113016. **[Published, SCIE, IF:5.452]**
3. Jain, L., Katarya, R., & Sachdeva, S. (2020). Recognition of opinion leaders coalitions in online social network using game theory. Knowledge-Based Systems, 203, 106158. **[Published, SCI, IF:5.921]**
4. Jain, L., Katarya, R., & Sachdeva, S. Opinion Leaders for information diffusion using Graph Neural Network in Online Social Network. **[Communicated]**
5. Jain, L., Katarya, R., & Sachdeva, S. Impact of Opinion Leader to control Covid-19 rumor in Online Social Network. **[Communicated]**

International Conferences

1. Jain, L., & Katarya, R. (2018, February). A Systematic Survey of Opinion Leader in Online Social Network. In 2018 International Conference on Soft-computing and Network Security (ICSNS) (pp. 1-5). **IEEE**.
2. Jain, L., & Katarya, R. (2018, December). Identification of opinion leader in online social network using fuzzy trust system. In 2018 IEEE 8th International Advance Computing Conference (IACC) (pp. 233-239). **IEEE**.
3. Jain, L., Katarya, R., & Sachdeva, S. (2019, August). Role of Opinion Leader for the diffusion of products using Epidemic model in Online Social Network. In 2019 Twelfth International Conference on Contemporary Computing (IC3) (pp. 1-6). **IEEE**.

4. Jain, L., Katarya, R., & Sachdeva, S. (2019, November). Opinion Leader discovery based on text analysis in Online Social Network. In 2019 4th International Conference on Information Systems and Computer Networks (ISCON) (pp. 446-450). **IEEE**.
5. Jain, L., Katarya, R., & Sachdeva, S. (2021, April). Applicability of the Opinion Leader to spread COVID-19 vaccine awareness through Online Social Network based on Sentiment Analysis. In 5th International Conference on Advances in Computing and Data Sciences (ICACDS). **Springer**.

Award

Commendable Research Award-2020

Commendable Research Award-2021

Biography

(Research Scholar: 2017-21)

Mr. Lokesh Jain is currently designated as a Ph.D. research scholar in the Department of Computer Science, Delhi Technological University, Delhi, India. He has completed his M.Tech from NSIT, Delhi, in Information System from the Department of Computer Science. He has completed an undergraduate degree from UPTU with honor. He has published various research papers in SCI/IEEE/SCOPUS indexed International Conferences/Journals. He is also a reviewer of the SCOPUS indexed Journal. He has more than 14 years of teaching experience in the reputed institutes. His research area of interest includes Social Networking, Computational Intelligence, Machine learning, Data mining, Fuzzy Logic, Software Testing, and Object-Oriented Techniques. At present, he is doing his research on opinion leader detection in an online social network using computational intelligence techniques.

He is also awarded the eminent *"Commendable Research Award"* in 2020 and 2021, respectively, from DTU, Delhi.